

UC San Diego

UC San Diego Previously Published Works

Title

Coarse predictions of dipole reversals by low-dimensional modeling and data assimilation

Permalink

<https://escholarship.org/uc/item/3zn2c0tr>

Authors

Morzfeld, M
Fournier, A
Hulot, G

Publication Date

2017

DOI

10.1016/j.pepi.2016.10.007

Peer reviewed

Coarse predictions of dipole reversals by low-dimensional modeling and data assimilation

Matthias Morzfeld

*Department of Mathematics, University of Arizona, 617 N. Santa Rita Ave. P.O. Box
210089 Tucson, AZ 85721-0089 USA. Email: mmo@math.arizona.edu*

Alexandre Fournier, Gauthier Hulot

*Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot,
CNRS, 1 Rue Jussieu, 75005 Paris, France.*

Abstract

Low-dimensional models for Earth’s magnetic dipole may be a powerful tool for studying large-scale dipole dynamics over geological time scales, where direct numerical simulation remains challenging. We investigate the utility of several low-dimensional models by calibrating them against the signed relative paleointensity over the past 2 million years. Model calibrations are done by “data assimilation” which allows us to incorporate nonlinearity and uncertainty into the computations. We find that the data assimilation is successful, in the sense that a relative error is below 8% for all models and data sets we consider. The successful assimilation of paleomagnetic data into low-dimensional models suggests that, on millennium time scales, the occurrence of dipole reversals mainly depends on the large-scale behavior of the dipole field, and is rather independent of the detailed morphology of the field. This, in turn, suggests that large-scale dynamics of the dipole may be predictable for much longer periods than the detailed morphology of the field, which is predictable for about one century. We explore these ideas and introduce a concept of “coarse predictions”, along with a sound numerical framework for computing them, and a series of tests that can be applied to assess their quality. Our predictions make use of low-dimensional models and assimilation of paleomagnetic data and, therefore, rely on the assumption that currently available paleomagnetic data are sufficiently accurate, in particular with respect to the timing of reversals, to allow for coarse predictions of reversals. Under this assumption, we conclude that coarse predictions of

dipole reversals are within reach. Specifically, using low-dimensional models and data assimilation enables us to reliably predict a time-window of 4 kyr during which a reversal will occur, without being precise about the timing of the reversal. Indeed, our results lead us to forecast that no reversal of the Earth’s magnetic field is to be expected within the next few millennia. Moreover, we confirm that the precise timing of reversals is difficult to predict, and that reversal predictions based on intensity thresholds are unreliable, which highlights the value of our model based coarse predictions.

Keywords: Dipole reversal prediction, low-dimensional modeling, data assimilation, geomagnetic field variations

1. Introduction

Earth possesses a time-varying magnetic field which is generated and sustained against Ohmic decay by a fluid dynamo driven by convection in its interior. The geomagnetic field changes over a wide range of time scales, from years to millions of years, and its strongest component, the dipole, has the dramatic feature that it occasionally switches polarity, i.e. the geomagnetic North becomes South, and vice-versa (see, e.g., Hulot et al. (2010a)). Such reversals happened throughout the geological history of our planet and their occurrence is well documented over the past 150 million years (Cande and Kent, 1995; Lowrie and Kent, 2004). However, little is known about the mechanisms that lead to a reversal. For example, detailed changes in the geometry of the geomagnetic field during a reversal are still poorly documented, and the conditions under which the reversal is initiated in Earth’s core remain essentially unknown (see, e.g., Amit et al., 2010; Glatzmaier and Coe, 2015; Valet and Fournier, 2016, for recent reviews).

A direct approach to modeling the geomagnetic field is numerical simulation of rapidly rotating spherical fluid shells, such as Earth’s fluid outer core, where the dynamo is operating. The computational cost of this approach is large, in particular if one wants to study the dipole over geological time scales of millions of years, so that only investigations with relatively limited dynamo simulations could be used so far (see, e.g., Lhuillier et al., 2013; Olson et al., 2013; Wicht and Meduri, *subm*). An alternative to direct numerical modeling is low-dimensional modeling. The idea is to derive a simplified representation of the large scale dynamics of a complex system while neglecting smaller scales. Several low-dimensional models have already been proposed for cloud

26 modeling (Koren and Feingold, 2011; Feingold and Koren, 2013) and for mod-
 27 eling of Earth’s dipole (Rikitake, 1958; Nozières, 1978; Hoyng et al., 2001;
 28 Brendel et al., 2007; Pétrélis and Fauve, 2008; Pétrélis et al., 2009; Kuipers
 29 et al., 2009; Gissinger et al., 2010; Gissinger, 2012; Buffett et al., 2013, 2014;
 30 Buffett and Matsui, 2015; Buffett, 2015; Meduri and Wicht, 2016). In the
 31 context of Earth’s magnetic field, a low-dimensional model represents the
 32 effects of complex interaction of the magnetic field and fluid flow, however
 33 the details of these interactions are not resolved. The heuristic arguments
 34 for the validity of these models are that the magnetic diffusivity is larger
 35 than the kinematic viscosity, which implies that the small scale magnetic
 36 field, induced by small scale velocity modes, is strongly damped, and, thus,
 37 the dynamics are dominated by a few magnetic modes (Gissinger, 2012).
 38 However, work that investigates the “usefulness” of low-dimensional models
 39 quantitatively is still missing. Here, “useful” is to be understood in the sense
 40 that low-dimensional models can reproduce paleomagnetic data, and that
 41 the models produce reliable predictions of large scale dynamics. Indeed, one
 42 of the main goals of this paper is to establish a suitable set of tests that can
 43 be used to quantify the utility of low-dimensional models for the geodynamo.

44 We present a data-driven, Bayesian approach and we calibrate the models
 45 against paleomagnetic data by “data assimilation”, i.e., we estimate model
 46 states from data by Bayesian statistics (see, e.g., Chorin and Hald (2013)).
 47 The data are the signed relative paleointensities which provide estimates
 48 of the strength of the axial dipole and its polarity over the past 2 Myr.
 49 The relative paleointensity is provided by Sint-2000 (Valet et al., 2005) and
 50 PADM2M (Ziegler et al., 2011) data sets, the polarity can be derived from the
 51 geomagnetic polarity time scale (Cande and Kent, 1995; Lowrie and Kent,
 52 2004). We consider four low-dimensional models:

- 53 (i) the deterministic three-variable model presented in Gissinger (2012),
 54 which we call G12;
- 55 (ii) the stochastic model presented in Buffett et al. (2013), which we refer
 56 to as B13;
- 57 (iii) the stochastic model derived in Pétrélis et al. (2009), which we abbreviate by P09;
- 58 (iv) a new scalar stochastic model that combines the numerical techniques
 59 used in Buffett et al. (2013) with the G12 model; we call this model the
 60 G12 based SDE.

62 Our data assimilation results (section 3) indeed establish compatibility of
 63 models and data in the sense that an average error after assimilation is no
 64 larger than 8% for all models and data sets we tried, provided that suitable

numerical techniques are used. This result is robust to variations in how the data are assimilated or how the data were obtained, since we obtain quantitatively similar results with several numerical data assimilation methods (see appendix Appendix B) and with both data sets.

The compatibility of low-dimensional models and paleomagnetic data suggests that general conditions for reversals to occur mainly result from the large-scale behavior of the dipole field, with the detailed morphology of the field playing a role only once such general conditions are met. If this were indeed the case, one could predict the large-scale dipole field over long time-scales, perhaps several thousand years. We investigate this possibility in section 4 where we introduce the concept of “coarse predictions” for dipole reversals. Specifically, we determine if we can identify *time-windows* of a few millennia during which reversals are likely to occur, without being precise about the timing of reversals within the time-windows. The temporal horizon of our predictions is comparable to the time needed for a reversal to occur, but shorter than the typical time elapsed between reversals. Coarse predictions could thus provide an “early warning system”, indicating that a reversal might occur within the next few millennia.

We present a series of tests to investigate if our proposed framework, which relies on low-dimensional models and data assimilation, produces more reliable predictions than several purely data-based prediction strategies. Predictions obtained in this way rely on the assumption that the paleomagnetic data, as documented by Sint-2000 and PADM2M, (one data point every 1,000 years) are sufficiently accurate for this purpose. Conditional on the latter assumption, we conclude that coarse predictions are indeed within reach, even with simple low-dimensional models. This highlights the value of low-dimensional models and data assimilation as an effective tool for addressing questions that are difficult to answer by other techniques, in particular direct numerical modeling. Perhaps more importantly, the coarse predictions we present, and the series of tests we suggest, may be useful to assess the utility of a future generations of improved low- or “intermediate”-dimensional models.

2. Paleomagnetic data and low-dimensional models

2.1. Paleomagnetic data

The data we use are the signed relative paleointensity of the past 2 Myr. These intensities describe estimates of the strength of the axial dipole, and

101 are available in the Sint-2000 (Valet et al., 2005) and PADM2M (Ziegler
 102 et al., 2011) data sets with a 1 kyr time step. The polarity is encoded by
 103 the sign of the dipole, which is taken from the geomagnetic polarity time
 104 scale (Cande and Kent, 1995; Lowrie and Kent, 2004). To find the exact
 105 timing of the polarity changes we proceed in slightly different ways for Sint-
 106 2000 and PADM2M. In the case of Sint-2000, we assume that reversals occur
 107 at time of polarity changes as confirmed from inspection of the original direc-
 108 tional information of Valet et al. (2005) (J.P. Valet, personal communication).
 109 In the case of PADM2M, however, we do not have access to analogous di-
 110 rectional information. We therefore a priori assumed the same timing as for
 111 Sint-2000, and checked that reversals did correspond to a minimum in the
 112 intensity record provided by PADM2M to within 1kyr. This turned out to be
 113 the case for most reversals, except for the Bruhnes Matuyama reversal and
 114 the two reversals bounding the Cobb mountain subchron. For these three
 115 reversals, a slight time shift was introduced to reconcile their timing with
 116 that of intensity lows in PADM2M, resulting in slight shifts in the timing of
 117 the sign changes in the PADM2M signed relative paleointensity with respect
 118 to that of Sint-2000.

119 For each data set, a unit relative paleointensity corresponds to a virtual
 120 axial dipole moment of $7.46 \cdot 10^{22} \text{ Am}^2$, as in Valet et al. (2005). Both data
 121 sets contain the relative paleointensity along with a Gaussian error model,
 122 i.e., every 1 kyr a datum of the paleointensity is available along with an es-
 123 timated standard deviation. However, the standard deviations of PADM2M
 124 are significantly smaller than those of Sint-2000. While the small errors of
 125 PADM2M may be accurate representations of the “pure” data error, they
 126 seem unreasonably small in the context of data assimilation. The reason is
 127 that these errors must describe a combination of “measurement errors”, i.e.,
 128 the uncertainty of the data, and “model errors”, i.e., how good the (low-
 129 dimensional) model is. We thus adjust the errors in PADM2M to account
 130 for model error. In the data assimilation (see section 3) we use the Sint-2000
 131 standard deviations for the PADM2M data. In particular, we find that the
 132 data assimilation is more stable with the larger standard deviations of Sint-
 133 2000. Figure 1 shows the mean and 95% confidence interval of the Sint-2000
 134 data as well as the mean of PADM2M.

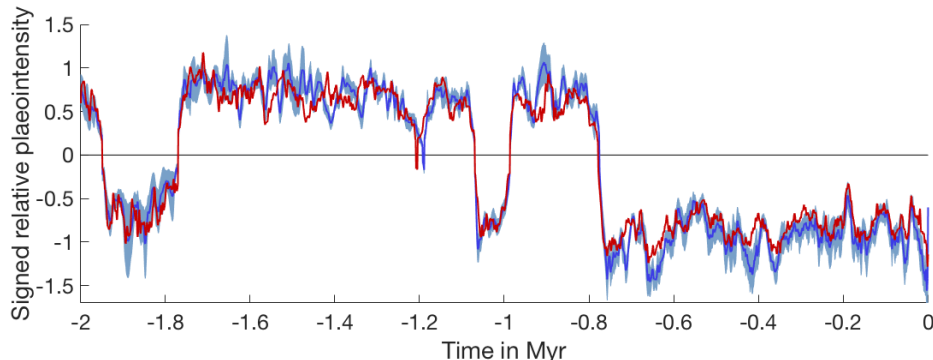


Figure 1: Signed relative paleointensity. The blue line represents the signed Sint-2000 data (Valet et al., 2005) and the light blue cloud represents a 95% confidence interval. The red line represents mean of the PADM2M data (Ziegler et al., 2011).

135 2.2. Scalar stochastic differential equation models: P09 and B13

136 The P09 (Pétreilis et al., 2009) and B13 (Buffett et al., 2013) models are
 137 stochastic differential equations (SDE) of the form

$$dx = f(x)dt + g(x)dW, \quad (1)$$

138 where the state, x , is either directly or indirectly related to the geomagnetic
 139 dipole, $f(x)$ and $g(x)$ are scalar functions, W is a Brownian motion, and t
 140 is time. A Brownian motion has the characteristics that it is almost surely
 141 continuous everywhere, that increments are independent Gaussian random
 142 variables $W(t) - W(s) \sim \mathcal{N}(0, t - s)$, and that $W(0) = 0$. Here and below,
 143 $\mathcal{N}(\mu, \sigma^2)$ is our notation for a Gaussian random variable with mean μ and
 144 variance σ^2 . The two models differ in their functions $f(x)$ and $g(x)$ and in
 145 the way x is related to the geomagnetic dipole.

146 The B13 model (Buffett et al., 2013) postulates that the dipole dynam-
 147 ics are governed by an SDE of the form (1), for which the state x is the
 148 geomagnetic dipole, and where the Brownian motion describes the effects of
 149 turbulent fluctuations of a velocity field. The drift and diffusion coefficients,
 150 $f(x)$ and $g(x)$, are estimated from paleomagnetic data. Specifically, the drift
 151 is derived from a double-well potential, i.e., Earth's dipole is modeled by a
 152 particle in a double-well, where each well represents a polarity. The particle,
 153 located in one of the wells, gets pushed around by noise, and the effects of the

154 noise may push the particle to overcome the potential barrier, thus complet-
 155 ing a reversal of the dipole. In Buffett et al. (2013), the drift and diffusion
 156 coefficients are estimated from Sint-2000 and PADM2M. Below we use the
 157 one resulting from PADM2M, and refer to Buffett et al. (2013) for the details
 158 of the numerics and their tuning. Since the drift and diffusion parameters
 159 are estimated from paleomagnetic data, the variable t of the resulting SDE
 160 model is “automatically” scaled as time. A typical simulation with B13 is
 161 shown in the upper-left panel of figure 2.

162 The B13 model has been used in other contexts as well. In Buffett et al.
 163 (2014), the same stochastic modeling approach was applied to data from nu-
 164 merical dynamo models, and in Buffett and Matsui (2015), the stochastic
 165 term of the B13 model was modified to account for correlations in time. Buf-
 166 fett (2015) used yet another variant of this model to study reversal duration
 167 and the intensity of fluctuations during a reversal. A model similar to the
 168 B13 model has also been discussed by Hoyng et al. (2001), and later by Bren-
 169 del et al. (2007) and Kuipers et al. (2009), who relied on a different numerical
 170 method to estimate the drift and diffusion coefficients. However, the details
 171 of how the drift and diffusion coefficients are computed are not important
 172 for our purposes. Finally, we note that the B13 model was recently revisited
 173 by Meduri and Wicht (2016), who relied on numerical dynamo simulations
 174 and paleomagnetic data to build SDE models of the form (1).

175 The P09 model (Pétronis et al., 2009) is based on the assumption that a
 176 general mechanism for field reversals exists, and that this process is largely in-
 177 dependent of the details of the velocity field. Specifically, the model describes
 178 the interaction of two modes of comparable thresholds, i.e., the magnetic field
 179 is $B(r, t) = a(t)B_1(r) + b(t)B_2(r)$. By imposing the symmetry of the equa-
 180 tions of magnetohydrodynamics $B \rightarrow -B$ in the amplitude equation, and by
 181 assuming that the amplitude has a shorter time scale than the phase, one
 182 obtains an SDE for the phase of the form (1) with

$$f(x) = \alpha_0 + \alpha_1 \sin(2x), \quad g(x) = 0.2\sqrt{|\alpha_1|}. \quad (2)$$

183 The dipole can be calculated from this phase by $D = R \cos(x + x_0)$. We
 184 use the same parameters as in Pétronis et al. (2009), $\alpha_1 = -185 \text{ Myr}^{-1}$,
 185 $\alpha_0/\alpha_1 = -0.9$, $x_0 = 0.3$. This choice of parameters also defines a time-scale
 186 for the variable t . Regarding the amplitude of the dipole, we set $R = 1.3$ to
 187 scale the P09 model output to have the same average relative paleointensity
 188 as the unsigned Sint-2000 data. With these parameters, the model exhibits

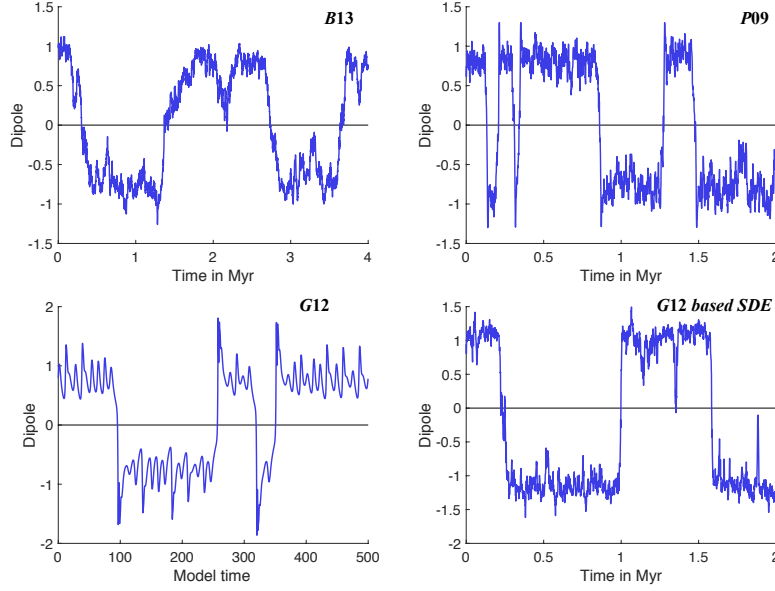


Figure 2: Dipole simulations with low-dimensional models. Top row: B13 (left) and P09 (right). Bottom row: G12 (left) and G12 based SDE (right). “Time” in the bottom row is dimensionless.

189 abrupt reversals and large fluctuations, as shown in the upper-right panel of
 190 figure 2, where a typical simulation result of P09 is shown.

191 The mechanism for reversals in the P09 model is as follows. The model
 192 has four fixed points, two are stable, and two are unstable. The two stable
 193 fixed points represent the two dipole polarities (North-South/South-North).
 194 The system hovers around one of the stable fixed points and gets pushed
 195 around by the noise (the Brownian motion), which represents the effects
 196 of turbulent fluctuations. When the deviation from the stable fixed point
 197 becomes large, the state can move beyond the neighboring unstable fixed
 198 point and then is attracted by the opposite stable fixed point, and a reversal
 199 of the dipole is completed. A more detailed discussion of the rich dynamics
 200 of this system is given in Pétrélis et al. (2009).

201 *2.3. The deterministic G12 model and the G12 based SDE model*

202 The G12 model consists of three deterministic ordinary differential equa-
 203 tions (ODE),

$$\frac{dQ}{dt} = \mu Q - VD, \quad \frac{dD}{dt} = -\nu D + VQ, \quad \frac{dV}{dt} = \Gamma - V + QD, \quad (3)$$

204 where $t \geq 0$ is to be identified as time, and where μ, ν and Γ are scalar pa-
 205 rameters, see Gissinger (2012). In this model Q represents the quadrupole,
 206 which may play an important role during reversals (McFadden et al., 1991;
 207 Glatzmaier and Roberts, 1995), D is the dipole and V represents the flow.
 208 The rich dynamics of these equations are studied by Gissinger (2012). In
 209 particular, it is shown that reversals are generated by crisis-induced inter-
 210 mittency when $\mu = 0.119$, $\nu = 0.1$, and $\Gamma = 0.9$ and that the model then
 211 shares a number of characteristics with the paleomagnetic data.

212 *2.3.1. Scaling of G12*

213 The G12 model is not equipped with a natural scaling of the amplitude
 214 of the dipole variable D to the geomagnetic dipole amplitude, or with a
 215 scaling of G12 model time, t , to geophysical time. To find the amplitude
 216 scaling of G12 we compute, as before, the average relative paleointensity
 217 of the unsigned Sint-2000 and PADM2M data sets and also compute the
 218 average of the absolute value of dipole variable of ten G12 model runs for
 219 250 dimensionless time units. By setting

$$\text{G12 amplitude scaling: } D = \sqrt{2} \times \text{relative paleointensity (signed),}$$

220 the average of the G12 dipole variable is approximately equal to the average
 221 relative paleointensity. Moreover, this scaling leads to good agreement of the
 222 histograms of the dipole variable D and of the signed relative paleointensity
 223 of Sint-2000 and PADM2M (left panel of figure 3). A typical simulation with
 224 G12 is shown in the lower left panel of figure 2.

225 To find the scaling of G12 model time, we may use the fact that the
 226 distribution of chron duration, i.e., the distribution of the time periods during
 227 which the geomagnetic dipole is in a stable polarity, is well approximated by
 228 a gamma distribution for both the paleomagnetic data (Lowrie and Kent,
 229 2004; Cande and Kent, 1995) and the G12 model, as shown by Gissinger
 230 (2012). By matching the shape parameters of a gamma distribution from

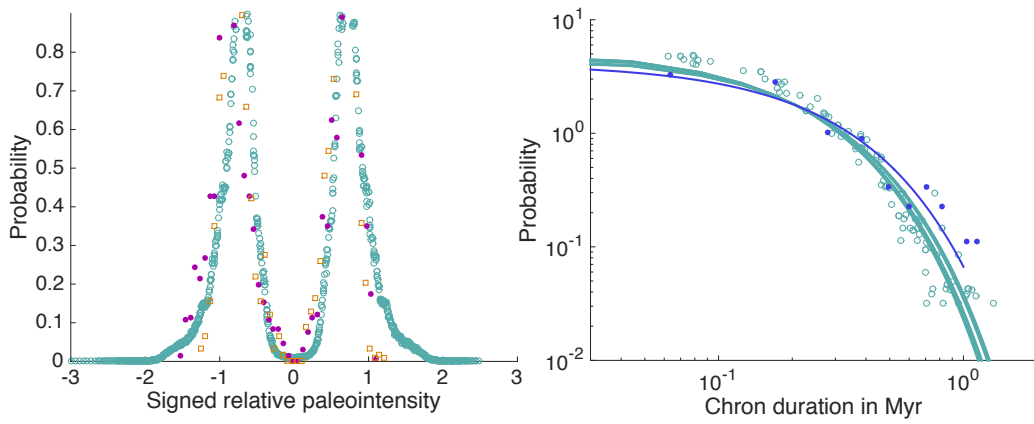


Figure 3: Left: histogram of signed relative paleointensity of Sint-2000 (purple dots), PADM2M (orange squares) and G12 after amplitude scaling (turquoise circles). Right: histogram of the chron duration over the past 30 Myr (blue dots, data from the CK95(1)(Cande and Kent, 1995) data set) and the maximum-likelihood gamma distribution fit (blue line); also shown are a histogram of G12 simulation data scaled using the geological time scale of 1 unit = 1kyr (turquoise circles) and maximum-likelihood fits for gamma distributions for each run (turquoise lines). See also figures 13 and 12 in Gissinger (2012).

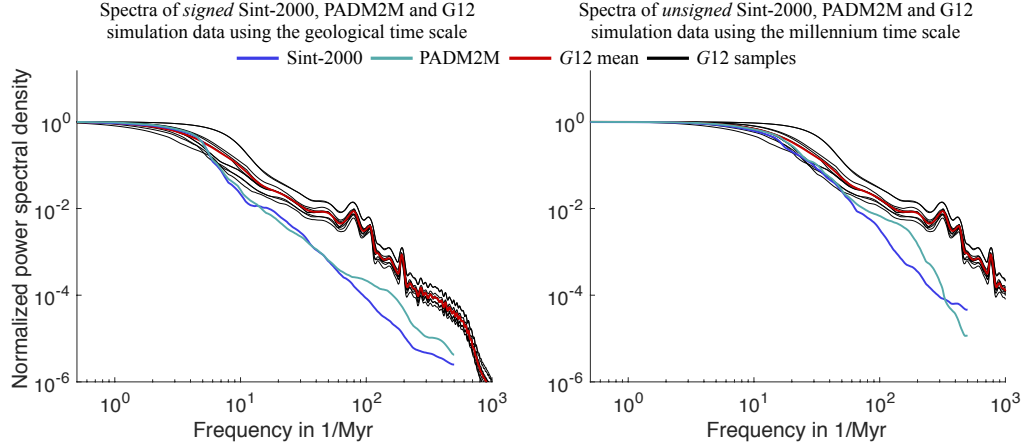


Figure 4: Left: spectra of Sint-2000 (blue), PADM2M (turquoise) and G12 (red, average of 10 model runs of 10^4 time units, each in black) when scaling to the geological time scale of 1 unit = 1kyr. Right: spectra of Sint-2000 (blue), PADM2M (turquoise) and G12 (red, average of 10 model runs of 10^4 time units, each in black) when scaling to the millennium time scale of 4 unit = 1kyr

G12 simulation data with the shape parameters of a gamma distribution of the paleomagnetic chron durations, we derive the

G12 geological time scale: 1 unit of G12 dimensionless model time = 1 kyr.

The shape parameters are computed by maximum likelihood estimation. For the paleomagnetic chron durations, these parameters are estimated from the CK95(1) data set of Cande and Kent (1995) as defined in Lowrie and Kent (2004), which contains the sign of the dipole over the past 30 Myr. For the G12 model, the parameters are estimated from ten simulation for 10^4 dimensionless time units. The right panel of figure 3 shows histograms and corresponding gamma distributions for CK95(1) and G12 when using this geological time scale.

It is instructive to assess this scaling by comparing the power spectral densities of G12 simulation data and Sint-2000/PADM2M data. We compute these spectra by the multi-taper spectral estimation technique described in Constable and Johnson (2005). The spectra are shown in the left panel of figure 4. Note that the first corner frequencies of the G12 model and of the Sint-2000 and PADM2M data match, but that the G12 model has a larger high-frequency content than PADM2M or Sint-2000 (by roughly

248 one order of magnitude for frequencies of 2 Myr^{-1} and above). We can
 249 attribute the low-frequencies to the occurrence of reversals, and the high
 250 frequencies to millennium scale dipole variations during chrons. This suggests
 251 that, when scaled using the above geological time scale, the dynamics of
 252 G12 essentially match the reversal statistics of the geomagnetic dipole, but
 253 fail to match its millennium behavior. We note that the high frequency
 254 content of Sint-2000 and PADM2M could be underestimated because the data
 255 are obtained by averaging over stacks, which possibly smoothes the signal.
 256 Indeed, Constable and Johnson (2005) constructed a spectral model whose
 257 high-frequency content is also larger than that of PADM2M or Sint-2000.
 258 However, we also note that the above geological time scale was computed
 259 using reversal statistics over the past 30 Myr, a period during which the
 260 reversal rate has increased by a factor of about 2 (see, e.g., Gallet and Hulot
 261 (1997)). A geological time scale estimated from more recent epochs would
 262 have been larger.

263 The mismatch of model and data for high-frequencies suggests that the
 264 geological time scale may not be optimal for scaling the G12 model, in par-
 265 ticular because the G12 model cannot be scaled to simultaneously match the
 266 geological and millennium dynamics of the Earth’s dipole field. This can be
 267 further illustrated by comparing spectra of *unsigned* data, shown in the right
 268 panel of figure 4. The low frequencies of the spectra of unsigned data are
 269 no longer dominated by reversal frequencies, with reversals occurring over
 270 millions of years, but are rather representative of field variations over mil-
 271 lennia. By comparing spectra of unsigned data, we find that matching the
 272 millennium scale of Earth’s dynamics to the “millennium” variation of the
 273 G12 model requires a time-scale four times larger than when matching model
 274 time to geological time scale. We thus define the

G12 millenium time scale: 1 unit of G12 dimensionless model time = 4 kyr.

275 In our attempts to assimilate data in the G12 model (see section 3), we
 276 observe that results improve dramatically when this millennium time scale is
 277 used, rather than the geological time scale, independently of the numerical
 278 data assimilation technique we use. This is an important observation. The
 279 reason is that reversals are rare, there are only 7 reversals within the 2000
 280 data points we consider. This implies that an accurately represented mil-
 281 lennium variation is more important for successful data assimilation than an
 282 accurate representation of the average time elapsed between reversals, i.e.,

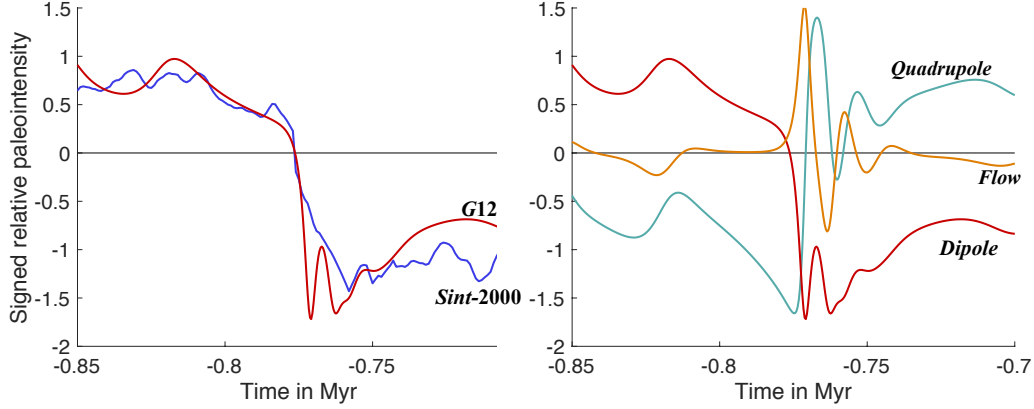


Figure 5: Left: Example reversal in a free run of the G12 model (red) scaled to the millennium time scale (1 unit = 4 kyr) and artificially synchronized to the time of the Brunhes-Matuyama reversal as seen in the Sint-2000 data (blue). Right: Behavior of the dipole (red), flow (yellow), quadrupole (turquoise) parameters of the G12 model during that same reversal (after figure 14 of Gissinger (2012)). The amplitude of the G12 model state variables D , Q , and V has been scaled by the factor $1/\sqrt{2}$.

the geological time scale. Indeed, with our millennium time scale, the G12 model encouragingly captures much of the behavior of the dipole before and during a reversal (left panel of figure 5). For the rest of this paper, we thus only use the millennium time scale.

Figure 5 illustrates the typical reversing behavior of the G12 model. We observe that the dipole slowly decreases and then quickly reverses, as is also observed in all reversals of the Sint-2000 data. The dipole reversal is followed by an overshoot, and such overshoots, perhaps less pronounced, are also observed in the data Valet et al. (2005). The right panel of figure 5 further illustrates the behavior of the flow and quadrupole variables during a dipole reversal. Specifically, when the dipole decreases, the quadrupole variable increases, and then reverses with the dipole. A strong peak can be observed in the velocity during a reversal.

Another dynamic time scale worth looking into is the e -folding time of the G12 model. This e -folding time is defined as the time it takes for the “distance” between two G12-trajectories to be multiplied by a factor e , and is an indicator of the intrinsic predictability of the G12 model. Its average value is estimated to be around 40 kyr (see Appendix Appendix A). This is much larger than the 30 year e -folding time found in three-dimensional

simulations, which must also account for the complex and fast-evolving non-dipole field (Hulot et al., 2010b; Lhuillier et al., 2011a). Provided that the G12 model can provide a useful coarse representation of the Earth’s dipole field with only three variables, this thus suggests that the G12 model could indeed be used to predict the average dipole field evolution over time-scales of several kyr. Such “coarse” predictions are precisely what we aim at. We investigate these ideas in more detail in section 4.

2.3.2. G12 based SDE

We further use the G12 model to propose an additional scalar SDE model, similar to the B13 model. We mimic the construction of the B13 model, but substitute the paleomagnetic data (Sint-2000 or PADM2M) with synthetic data from G12 scaled to the millennium scale as described above. In constructing a G12 based SDE model, we postulate an SDE (1) for the dipole of the G12 model and use the numerical techniques of Buffett et al. (2013) to estimate the drift and diffusion coefficients from G12 simulation data (rather than from paleomagnetic data). Specifically, we fit a cubic function to the drift and a quadratic function to the square root of the diffusion coefficient. We refer to this model as the “G12 based SDE”. A typical simulation with the G12 based SDE is shown in the lower right panel of figure 2.

3. Data assimilation results

We perform data assimilation using the various numerical methods described in appendix Appendix B and the two data sets Sint-2000 and PADM2M. For each model, data set and data assimilation technique, we compute the relative error of the assimilation over the 2 Myr period defined by

$$e = \frac{\sum_{n=1}^{2000} \left(z^n - \hat{E}[x^n | z^{1:n}] \right)^2}{\sum_{n=1}^{2000} (z^n)^2}, \quad (4)$$

where z^n are the data at time n kyr and $\hat{E}[x^n | z^{1:n}]$ is the approximation of the conditional mean of the dipole given the data up to time n kyr. The conditional mean is the minimum mean square error estimate of the state, see, e.g., Chorin and Hald (2013). Each method resorts to a finite number of model samples, also called particles, whose distribution aims at providing a faithful description of the model uncertainties. For each method, we vary the number of samples from 50 to 400, compute the above error, and check

		B13			G12 based SDE			P09	G12			
Method:		S-EnKF	SIR	S-IMP	S-EnKF	SIR	S-IMP	SIR	D-EnKF	D-IMP		
Data/sweep:		1	1	1	1	1	1	1	1	1	5	10
# samples												
Sint-2000	50	5.61	6.72	5.91	2.57	2.94	2.64	7.91	30.5	5.70	3.74	4.28
	100	5.46	6.37	5.76	2.31	2.64	2.53	7.29	30.7	5.43	3.80	4.20
	200	5.53	6.25	5.65	2.37	2.57	2.48	7.83	30.0	5.38	3.61	6.19
	400	5.46	6.10	5.63	2.32	2.51	2.51	7.35	29.5	5.39	3.51	6.18
PADM2M	50	5.37	8.03	5.39	2.15	2.47	2.22	9.06	27.1	6.63	5.09	5.98
	100	5.23	8.27	5.28	1.93	2.18	2.07	8.84	27.9	6.27	4.92	5.93
	200	5.23	7.51	5.27	1.72	2.08	1.91	8.94	26.5	5.99	4.99	5.93
	400	5.22	7.42	5.20	1.68	2.05	1.82	8.46	26.8	5.83	4.92	5.83

Table 1: Relative error (in %) of paleomagnetic data assimilation. We assimilate Sint-2000 and PADM2M into B13, G12 based SDE and P09 by S-EnKF, SIR and S-IMP, and into G12 by D-EnKF and D-IMP (see appendix Appendix B.2). We vary the number of samples to check that sampling error is not the dominating error and vary the number of data points used per assimilation sweep for G12 in D-IMP.

that sampling error is not the dominating error. In all cases, we observe that the error decreases when we increase the number of samples, but not by much, which indicates that 200-400 samples are sufficient to compute reliable estimates by Monte Carlo.

3.1. Data assimilation with scalar SDE models

We first consider the three scalar SDE models B13, G12 based SDE, and P09. We apply the ensemble Kalman filter for stochastic models (S-EnKF), sequential data assimilation with implicit sampling (S-IMP) and sequential importance sampling with resampling (SIR) to these models (see section Appendix B.2 in appendix Appendix B for a brief description of each method). For the P09 model we only used the SIR method. The reason is that the P09 model is “more nonlinear” than the B13 or G12 based SDE models, which makes the implementation of the other techniques more difficult. However, EnKF and S-IMP are techniques to keep the computational requirements of data assimilation reasonable and, since computation is not an issue here, using SIR is feasible. In each method, we use one observation per assimilation sweep. The results are listed in table 1. A typical result of data assimilation by the G12 based SDE and P09 model are shown in the left and right panels of figure 6. A typical result obtained by B13 is qualitatively similar.

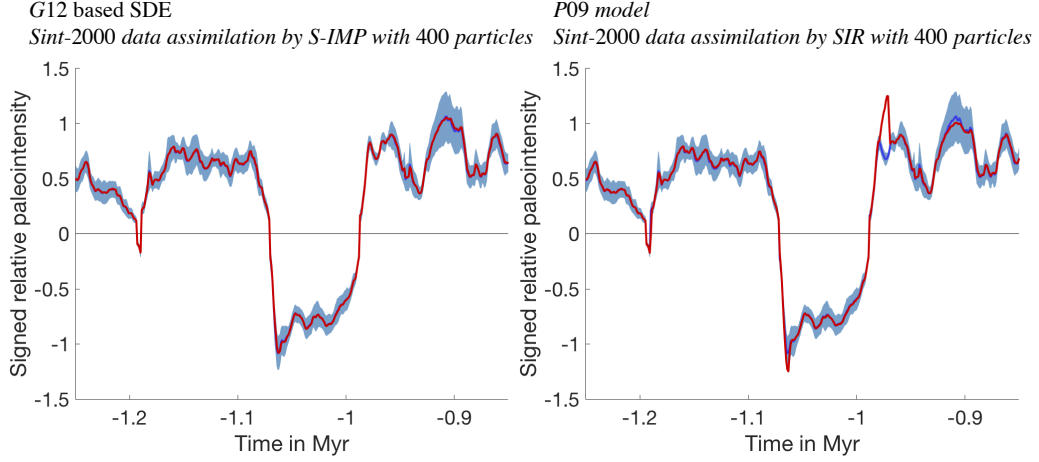


Figure 6: Assimilation of Sint-2000 data into G12 based SDE by the S-IMP method with 400 samples (left) and assimilation of Sint-2000 data into P09 model by the SIR method with 400 samples (right). Blue: Sint-2000 data. Light blue cloud: Sint-2000 data 95% confidence interval. Red: conditional mean obtained through the assimilation process.

352 For the B13 and G12 based SDE models, and using suitable numerical
 353 data assimilation, both data sets lead to errors no larger than 6%. Errors are
 354 only slightly larger in the case of the P09 model. Such small errors suggest
 355 that the “free dynamics” of the scalar models are, in principle, compatible
 356 with that of the geomagnetic dipole, in the sense that data assimilation can
 357 keep the model trajectories close to the data.

358 This positive result is perhaps not surprising for B13 and P09, because
 359 the parameters of these models are adjusted to match paleomagnetic data.
 360 Specifically, drift and diffusion coefficients of the B13 model are estimated
 361 from the PADM2M data we assimilate, and the model parameters of the P09
 362 model are chosen to “fit paleomagnetic data” (Pétrellis et al., 2009). However,
 363 the parameters that define the G12 based SDE are not estimated from these
 364 data. Rather, the drift and diffusion coefficients that define the G12 based
 365 SDE model are estimated from “synthetic data” of the G12 model (with
 366 model-time appropriately scaled, see above). The small errors we obtain
 367 with the G12 based SDE thus imply that G12 itself may be compatible with
 368 the paleomagnetic data. We study this in more detail below.

3.2. Data assimilation with G12

We now consider the deterministic G12 model and use the EnKF for deterministic models (D-EnKF) and sequential data assimilation with implicit sampling for deterministic models (D-IMP) to assimilate the PADM2M and Sint-2000 data. The D-EnKF and D-IMP techniques are described in detail in section Appendix B.1 of appendix Appendix B. When considering sequential data assimilation with implicit sampling, we can vary the number of data points we assimilate per sweep (see appendix Appendix B.1.2). Specifically, one can attempt to assimilate the 2 Myr of data in one sweep, i.e., one can try to find initial conditions for G12 that lead to a trajectory of the dipole variable that is compatible with the paleomagnetic data. However, this approach did not prove successful because the optimization required for implicit sampling failed to converge. The reasons for this failure are that (i) the G12 model cannot account simultaneously for the millennium and geological time scales of dipole fluctuations, whereas an assimilation over 2 Myr of data in one sweep assumes that both time scales are correctly represented (see section 2.3.1); and (ii) the e -folding time of the G12 model of about 40 kyr makes it numerically difficult to propagate information from data backwards over several million years. To address these difficulties, we apply data assimilation sequentially as described in appendix Appendix B.1.2. Specifically, we assimilate 1-15 kyr of data per sweep. The results are shown in table 1. A typical result of data assimilation with G12 is shown in the top-left panel of figure 7. We observe that we obtain similar errors when assimilating 1 or 5 data points per sweep, however the assimilation result is a lot smoother when we use 5 data points per sweep. We further observe that the error increases steeply if more than 5 kyr of data are assimilated per sweep. Further, we observe that EnKF yields a larger error than implicit sampling. The reason may be that G12 is more nonlinear than the B13 model or the G12 based SDE model, in particular due to the Q and V variables. This makes the use of a nonlinear data assimilation method more important, because the Gaussian approximation of EnKF may not be valid.

It is evident from figure 7, that significant discontinuities occur at each time we assimilate data, i.e., every 5 kyr. These discontinuities indicate that assimilating the next 5 kyr of data has a large effect on the state estimate. This could be due to either an intrinsic incompatibility of the G12 model with the data, or large errors in the unobserved quadrupole and flow variables. We investigate this issue by using synthetic data shown in figure 8, generated as follows. We simulate the G12 model starting from initial conditions that

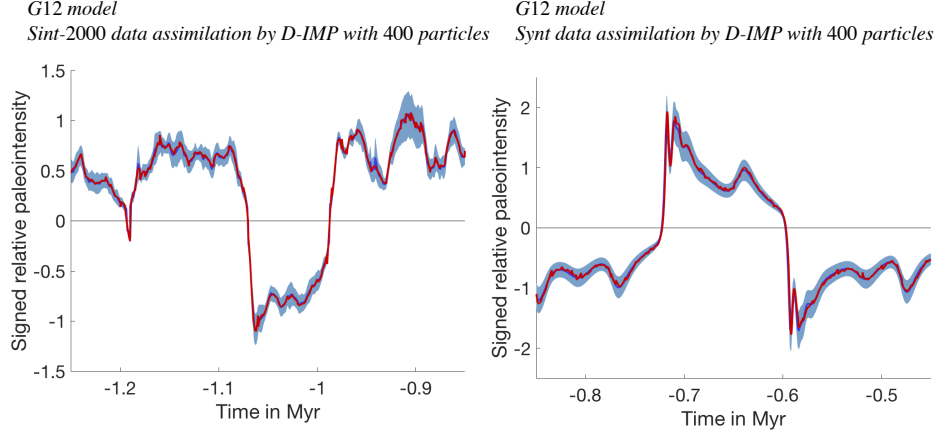


Figure 7: Result of data assimilation (red) and data (blue, often hidden), along with assumed errors in the data (light blue cloud). Left: Sint-2000 data. Right: synthetic data (Synt, see text and figure 8). Data assimilation is done by D-IMP with 400 samples, and 5 data points per sweep. Left: result of sequential data assimilation with D-IMP for G12 and using Sint-2000 data. Right: Same but when assimilating synthetic data (Synt, see text and figure 8).

lead to a dipole sequence similar to that of the paleomagnetic data. We record the state every 1 kyr over a 2 Myr period, and add random errors that are distributed similarly to those of Sint-2000. Specifically, the errors are Gaussian and the standard deviation is chosen such that the mean of the relative paleointensity divided by the standard deviation of the errors is the same for Sint-2000 and the synthetic data. For the rest of this paper, we will refer to this synthetic data set as the “Synt” data set.

We observe discontinuities when assimilating this synthetic data, as illustrated in the top-right panel of figure 7. This is an important observation, since, by construction, the Synt data are intrinsically compatible with the G12 model. Our numerical experiment thus indicates that the discontinuities observed when assimilating the paleomagnetic data are more likely to be caused by the assimilation method, and in particular by the fact that only dipole data are assimilated. Specifically, we find that the errors after assimilation in the unobserved Q and V variables are larger than the errors in the observed dipole variable, namely 20% error in Q , 51% error in V .

In summary, we obtain small errors of about 3-8% in the dipole variables of all models, provided an appropriate data assimilation technique and a

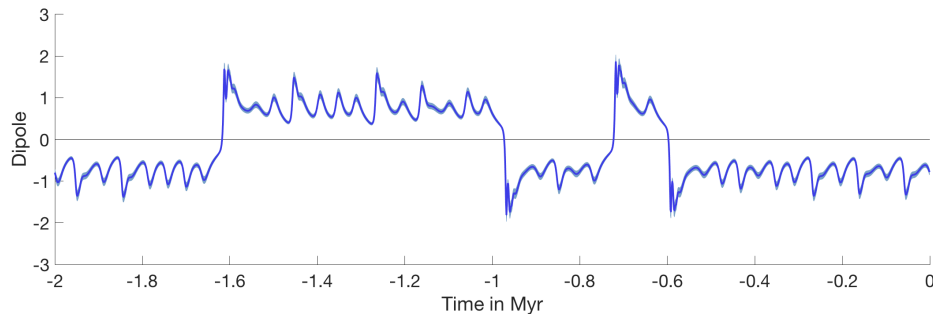


Figure 8: “Synt” synthetic data computed from the G12 model. The blue line represents the mean and the light blue cloud represents the 95% confidence interval.

modest number of data points per sweep are used. The small errors suggest that G12 is indeed compatible with the paleomagnetic data. As before, compatible means that the data assimilation can keep the G12 dipole variable close to the data. This result is conditional on that no more than 5 kyr of data are assimilated, so that the limitations of G12, discussed in section 2.3.1, do not come into play.

4. Coarse predictions of dipole reversals

In the above section we showed that four low-dimensional models could be calibrated to paleomagnetic data as documented by Sint-2000 and PADM2M, in the sense that the average error in (4) is below 8%. This suggests that using these models and data assimilation may lead to simplified yet useful representations of the Earth’s dipole, and that successful dipole reversal predictions may be based on calibrated model states. We investigate these issues carefully.

We wish to find out if low-dimensional models can reliably predict a time-window during which a reversal is likely to happen, without being precise about the timing of the reversal. The idea of such coarse prediction strategies is as follows. Given the model and data, a Monte Carlo based data assimilation computes a collection of model states that are compatible with the paleomagnetic data, in the sense that these states are samples from an appropriate posterior distribution. Each model state can be used to make a prediction by using it as an initial condition for a simulation over a specified time-window, called the “horizon”. This leads to a cloud of trajectories that

448 extend into the future, and these trajectories can be used to approximate
 449 the probability of a reversal within the horizon by computing the ratio of the
 450 number of trajectories that reverse to the total number of trajectories. For
 451 short horizons, the strategy “*predict that no reversal will occur within the*
 452 *horizon*” can be expected to be successful, and for extremely long horizons, a
 453 reversal becomes likely. We consider 4 kyr and 8 kyr horizons, because they
 454 are relevant, since the horizon is comparable to the time the system needs to
 455 reverse, but shorter than a typical chron.

456 4.1. *Hindcasting paleomagnetic data*

457 We assess the success of coarse predictions by “hindcasting”, i.e., by pre-
 458 dicting the past. This technique is routinely used in numerical weather pre-
 459 diction and goes as follows. One assimilates data up to a specified time in
 460 the past and computes model states that are compatible with the data up to
 461 that time. One then evolves each state by the model, without assimilating
 462 more data. The trajectories one obtains in this way “predict” what happened
 463 in the past. Thus, hindcasting assesses how successful a prediction strategy
 464 is for predicting the future, by testing how successful it performs for past
 465 events.

466 For the hindcasts illustrated in figures 9, 10, 11, and 12, we assimilate
 467 Sint-2000 data, however similar results are obtained when PADM2M is used
 468 for assimilation. The assimilation is done by D-IMP and 200 samples, 5 data
 469 points per sweep for the G12 model, by S-EnKF with 400 samples for the
 470 G12 based SDE model, by S-IMP with 400 samples for B13, and by SIR with
 471 400 samples for P09, for the reasons outlined in section 3.

472 We start by considering scalar SDE models, a typical example of which is
 473 the P09 model. In figure 9 we show P09 based hindcasts for a 4kyr horizon
 474 for the Brunhes-Matuyama (BM) reversal, which occurred between 777 and
 475 776 kyr ago. Before the BM reversal, at $t = -781$ kyr, the system appears to
 476 be close to a branching point as a significant number of samples tend towards
 477 a reversal, while the majority of the samples indicate that the dipole variable
 478 will increase (top-left panel). Only a few of the 400 samples exhibit a reversal
 479 within the horizon, so that the predicted probability of a reversal is small
 480 (7%). At $t = -777$ kyr, as the system gets closer to the BM reversal, the
 481 majority of samples aligns and exhibits a decrease in the dipole amplitude
 482 (top-right panel), with 40% of the samples exhibiting a reversal within 4 kyr.
 483 Note that the geomagnetic dipole indeed reverses during this time window,
 484 i.e., the BM reversal is correctly predicted by 40% of the P09 trajectories.

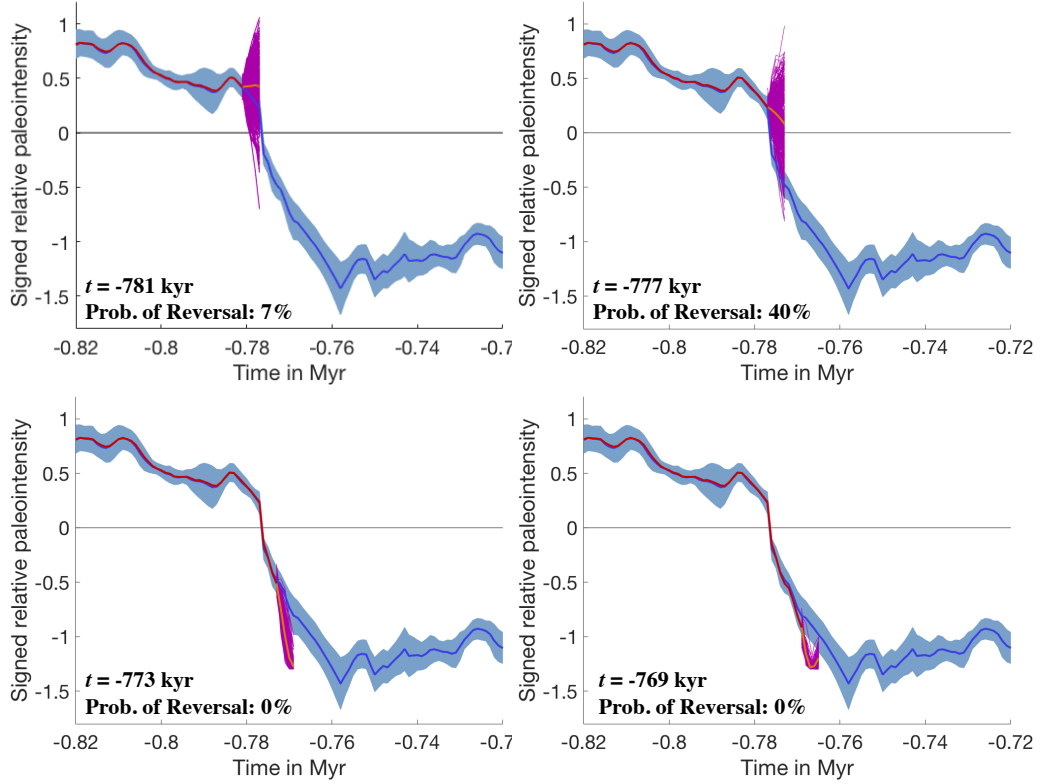


Figure 9: Hindcasting the Brunhes-Matuyama reversal by P09. Blue: Sint-2000 data. Light blue cloud: 95% confidence interval. Red: data assimilation (Sint-2000 data, SIR, 400 samples). Purple: predictions over 4 kyr. Orange: average of predictions over 4 kyr. Top left to bottom right: hindcasting starts at $t = -781$ kyr, $t = -777$ kyr, $t = -773$ kyr, $t = -769$ kyr.

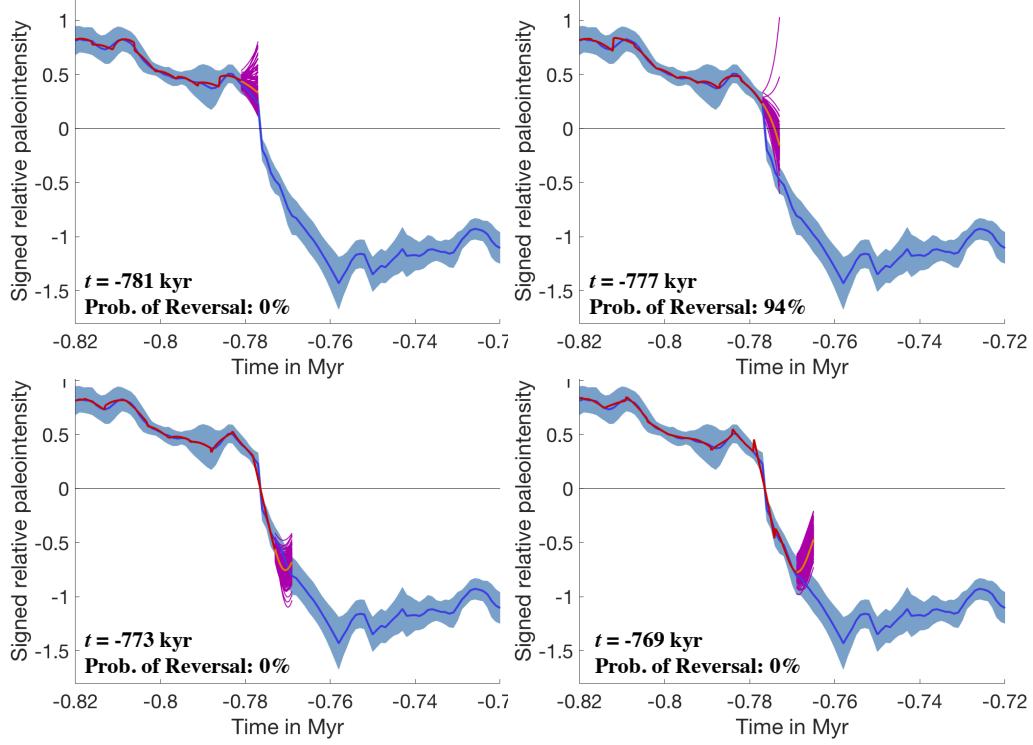


Figure 10: Hindcasting the Brunhes-Matuyama reversal by G12. Blue: Sint-2000 data. Light blue cloud: 95% confidence interval. Red: data assimilation (Sint-2000 data, D-IMP, 200 samples, 5 observations per sweep). Purple: predictions over 4 kyr. Orange: average of predictions over 4 kyr. Top left to bottom right: hindcasting starts at $t = -781$ kyr, $t = -777$ kyr, $t = -773$ kyr, $t = -769$ kyr.

At $t = -773$ kyr, all trajectories exhibit a quick decrease of the dipole, however the decrease is quicker than the data (bottom-left panel). At $t = -769$ kyr, all P09 trajectories exhibit an overshoot (bottom-right panel). An overshoot is also observed in the data, however the overshoot happens later than predicted by P09.

We now turn to the case of the deterministic G12 model, and show, for comparison, G12 based hindcasts of the BM reversal (Figure 10). We observe qualitatively similar results as when hindcasting by P09 (top row). However, the reversal is more accurately predicted by G12, since the majority of samples correctly predict that the dipole will decrease during the 4 kyr following $t = -781$ kyr. In fact, 94% of the trajectories reverse within 4

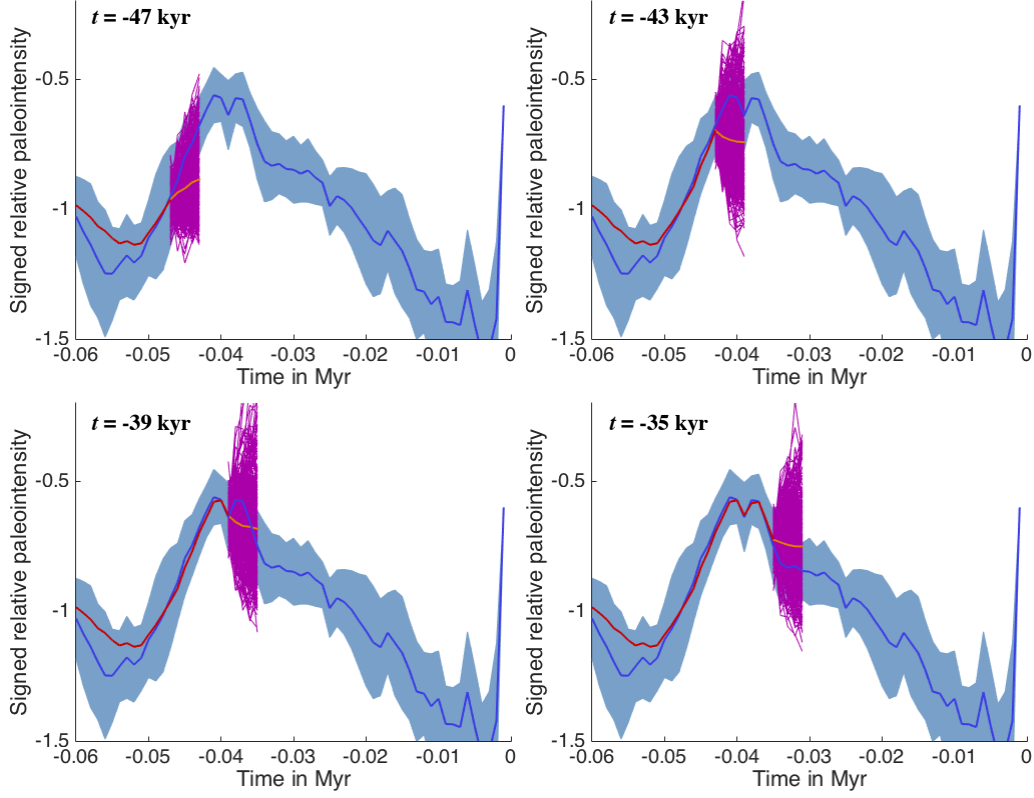


Figure 11: Hindcasting the Laschamp event by P09. Blue: Sint-2000 data. Light blue cloud: 95% confidence interval. Red: data assimilation (Sint-2000 data, SIR, 400 samples). Purple: predictions. Orange: average of predictions. Top left to bottom right: hindcasting starts at $t = -47$ kyr, $t = -43$ kyr, $t = -39$ kyr, $t = -35$ kyr.

kyr of $t = -777$ kyr, which is the time window during which the reversal indeed occurred. The G12 hindcasts right after the reversal on the other hand appear unphysical, and increase after a brief decrease of the dipole (bottom row). We observe this unphysical behavior when hindcasting all reversals of the past 2 Myr.

Also of great interest are hindcasts based on the low-dimensional models for the Laschamp low-intensity event, which occurred approximately 40 kyr ago and did not lead to a reversal. In figure 11 we show P09 based hindcasts during this event. We observe that none of the samples reverse within 4 kyr, which shows that the model correctly predicts that no reversal should have

506 occurred. However, the system seems to be in a state of branching, because
 507 a large number of the samples predict that the signed relative paleointensity
 508 should keep increasing for the next 4 kyr, while at the same time, a large
 509 number of samples also predict that the signed relative paleointensity should
 510 decrease.

511 In figure 12 we show G12 based hindcasts of the same Laschamp event.
 512 The results we obtain are qualitatively similar, however immediately after
 513 the dipole field reaches its maximum value (at $t = -39$ kyr and $t = -35$
 514 kyr), the G12 trajectories spread out more quickly than the samples of P09.

515 Indeed, we can perform hindcasts every 1000 years for all four models
 516 we consider, and compute the probability of a reversal to occur within a
 517 given horizon as a function of time. The results for all four low-dimensional
 518 models for a 4 kyr horizon when assimilating Sint-2000 data are shown in
 519 figure 13. We note that the probability graphs of all four models “peak”
 520 when the dipole indeed reverses. However, the B13 model assigns a low
 521 probability to the event “*a reversal occurs within 4 kyr*” at all times, even
 522 when a reversal is about to happen, with the maximum probability being
 523 about 30%. The graphs of the other three models, P09, G12 and G12 based
 524 SDE, look qualitatively similar to each other, and are somewhat noisier than
 525 the graph obtained with B13. We obtain qualitatively and quantitatively
 526 similar results when PADM2M data are assimilated.

527 4.2. Inverse relative Brier score

528 The key question is: *which model leads to the most valuable predictions?*
 529 To answer this question, we need a quantitative assessment of the validity
 530 of predictions. A convenient tool for providing such an assessment is the
 531 Brier score, which uses hindcasts to measure the mean square error between
 532 computed probabilities and the actual outcome (Brier, 1950). This Brier
 533 score is defined by

$$b = \frac{1}{N} \sum_{j=1}^N (p_j - o_j)^2, \quad (5)$$

534 where N is the number of hindcasts one makes, p_j is the predicted probability
 535 of an event, and o_j is a variable that is one if the event happens, and zero if
 536 it does not happen. For our purposes, the event is “*a reversal occurs within*
 537 *the horizon*”, and $N = 2000$, i.e., we make hindcasts at each time we have a
 538 new datum between 2 Myr and 1 kyr ago.

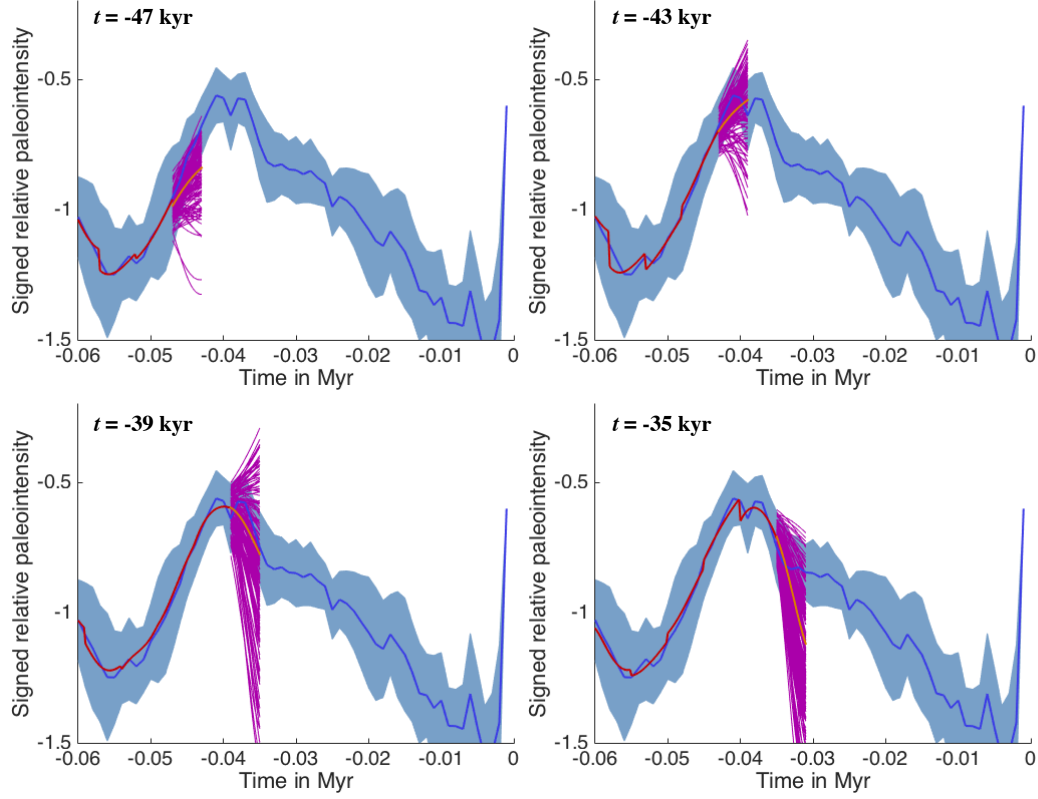


Figure 12: Hindcasting the Laschamp event by G12. Solid blue: Sint-2000 data. Light blue cloud: 95% confidence interval. Red: data assimilation (Sint-2000 data, S-IMP, 200 samples). Purple: predictions over 4 kyr. Orange: average of predictions over 4 kyr. Top left to bottom right: hindcasting starts at $t = -47$ kyr, $t = -43$ kyr, $t = -39$ kyr, $t = -35$ kyr.

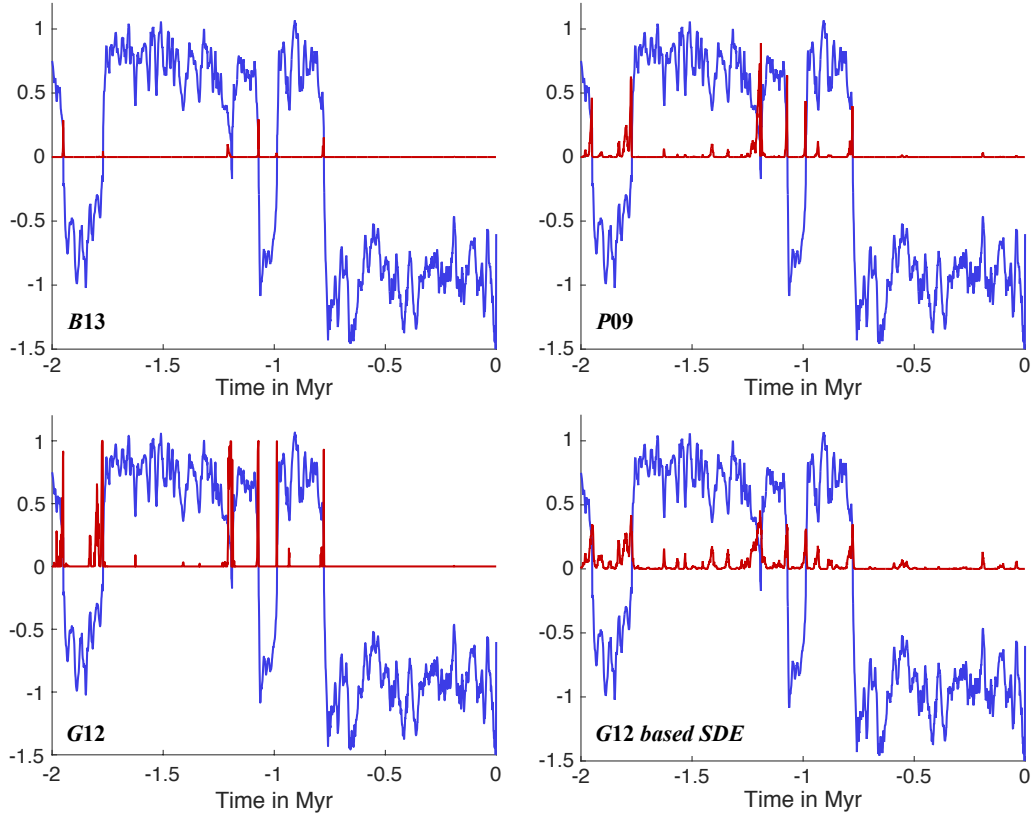


Figure 13: Hindcasting by low-dimensional models. Shown is the predicted probability of a reversal to occur within 4 kyr as function of time (red) along with the Sint-2000 data (blue). Top-left to bottom-right: B13, P09, G12 and G12 based SDE models.

539 We define a reference Brier score to assess how good coarse prediction
540 strategies perform. This reference Brier score relies only on reversal statistics.
541 Specifically, let R be the number of times the event “*a reversal happened*
542 *within the horizon*” happened, and let $N = 2000$ be the number of tries. The
543 probability that a reversal happens, based solely on the reversal statistics of
544 the past 2 Myr, is $p_{\text{stat}} = R/N$. For 4 kyr and 8 kyr horizons, $p_{\text{stat}} = 1.4\%$
545 and 2.6% , respectively. The reference Brier score can now be computed from
546 equation (5) by setting $p_j = p_{\text{stat}}$ for $j = 1, \dots, N$, with p_{stat} as above. For
547 the paleomagnetic data, the reference Brier scores are $b_{\text{ref}} = 0.013$ for a 4 kyr
548 horizon, and $b_{\text{ref}} = 0.025$ for a 8 kyr horizon. We define the inverse relative
549 Brier score (IRBS) as the ratio of the reference Brier score and the Brier
550 score of the prediction strategy we wish to asses:

$$\text{IRBS} = b_{\text{ref}}/b_{\text{model}}. \quad (6)$$

551 IRBS values larger than 1 thus indicate that the prediction strategy is on
552 average more reliable than a coin-toss, where the coin is biased by the prob-
553 ability p_{stat} . Note that such a coin does not at all behave like the “usual”
554 head-and-tails coin with probability $p_{\text{stat}} = 50\%$.

555 Below we use IRBS to quantify how reliable a prediction strategy is.
556 However, IRBS is far from being a perfect performance measure for dipole
557 reversal predictions. The reason is that the event “*no reversal occurs within*
558 *the horizon*” occurs more frequently than the event “*a reversal occurs within*
559 *the horizon*”. This means in particular that the strategy “*predict that no*
560 *reversal will ever happen*” scores an IRBS slightly larger than one (specifi-
561 cally, 1.01 for a 4 kyr horizon, 1.02 for a 8 kyr horizon). On the other hand,
562 this strategy is clearly not a good prediction strategy, since reversals are the
563 relevant events here. One should thus keep in mind that prediction strategies
564 that tend to assign a high probability to the event “*no reversal occurs within*
565 *the horizon*” might be rendered successful by our IRBS measure, despite the
566 fact they may grossly underestimate probabilities of reversals within time
567 windows when a reversal actually occurred. Inadequacy of IRBS is amplified
568 by limited amounts of data and these limitations are discussed further in
569 section 4.4 below.

570 4.3. IRBS comparison of data assimilation based prediction strategies

571 We compute IRBS for all four models, and when assimilating synthetic
572 and paleomagnetic data. Experiments with synthetic data are essential here

	Synthetic data		Sint-2000		PADM2M	
Horizon	4 kyr	8kyr	4 kyr	8kyr	4 kyr	8kyr
G12	3.49	0.47	1.21	0.35	1.13	0.25
G12 based SDE	1.40	1.43	1.28	1.23	1.19	1.10
P09	1.89	2.05	1.51	1.63	1.50	1.93
B13	1.42	1.41	1.01	1.15	1.06	1.08

Table 2: IRBS for G12, stochastic G12, P09, and B13 models for 4 kyr and 8 kyr horizons and using synthetic data, Sint-2000, and PADM2M. IRBS values above 1 indicate that the data assimilation based strategy has more predictive capability than guessing based on reversal statistics.

573 because these tests reveal whether or not the models are intrinsically pre-
574 dictable by the proposed strategy. Synthetic data are generated by the low-
575 dimensional models using the state trajectories already shown in figure 2.
576 Each data point has associated Gaussian errors whose variance is such that
577 the mean of the relative paleointensity divided by the standard deviation
578 of the errors is the same for Sint-2000 and for each of the four synthetic
579 data sets. As before, we consider 4 kyr and 8 kyr horizons. Our results are
580 summarized in table 2.

581 We find that all four models yield IRBS larger than one when synthetic
582 data are used and when the horizon is 4 kyr. This suggests that all models
583 are intrinsically predictable over a 4 kyr horizon by our proposed strategy.
584 We further obtain IRBS values larger than one for the B13, P09 and G12
585 based SDE models when considering predictions over a 8 kyr horizon. In
586 contrast, the G12 model yields an IRBS less than one, which suggests that
587 G12 is not intrinsically predictable over this longer horizon. The reason
588 could be large errors in the unobserved variables Q and V , which are proxies
589 for un-modeled field and flow components. Large errors in these variables
590 are indeed quickly amplified by G12’s dynamics, leading to trajectories that
591 spread out too quickly and too widely to be useful for predictions. In prin-
592 ciple “more accurate data”, or “more data”, i.e., data of the quadrupole and
593 velocity variables, could reduce these errors and make the G12 model pre-
594 dictable beyond the 4 kyr horizon, since its e -folding time is 40 kyr. In our
595 experiments, however, we have to adjust the synthetic data to have roughly
596 the same errors as the paleomagnetic data, and to acknowledge that data
597 of other field or flow components are not available at this point. Thus, our
598 synthetic data experiments suggest that, with the currently available paleo-

599 magnetic data, G12 can only predict dipole reversals within a 4 kyr horizon,
600 and not for longer horizons.

601 We observe a significant drop in IRBS for all models and considered hori-
602 zons when hindcasting paleomagnetic data. The reason is that model error
603 can be expected to be significant, since all models are simplified representa-
604 tions of Earth’s dipole dynamics. However, the results we obtain with either
605 paleomagnetic data set, Sint-2000 or PADM2M, are very similar and predic-
606 tions based on any of the four models still score IRBS larger than 1 for a 4
607 kyr horizon. P09, B13, and G12 based SDE also still score higher than 1 for
608 the 8 kyr horizon. In contrast, G12 scores below 1 for a 8 kyr horizon, as in
609 the above experiments with synthetic data.

610 Taken altogether, our assessment by IRBS is encouraging, as it suggests
611 that all models have some predictive power even when paleomagnetic data
612 are assimilated. In particular, we find that the P09 model scores the highest
613 IRBS. However, IRBS can be high for inadequate reasons, and, therefore, can
614 not represent sufficient evidence that a given prediction procedure is most
615 reliable. We therefore assess the model-based predictions by an additional
616 set of more stringent threshold-based prediction tests.

617 4.4. *Threshold-based predictions*

618 In threshold-based predictions one attaches a threshold to a parameter
619 of a dynamic system and determines the probability of an event to occur by
620 checking if the parameter is above or below the threshold. For example, one
621 assigns probability one, i.e, predicting with certainty that the event will oc-
622 cur, if the parameter is above the threshold, and one assigns probability zero,
623 i.e., predicting with certainty that the event will *not* occur, if the parameters
624 is below its threshold. Alternatively, one can assign probability one if the
625 parameter is below the threshold, and probability zero otherwise.

626 The success of threshold-based strategies depends on how the threshold
627 is chosen and below we use an objective way to do this by splitting available
628 data into two parts, “training data” and “verification data”. We first “learn”
629 the threshold from the training data as follows. We vary the threshold value,
630 infer the corresponding (zero or one) threshold-based probabilities at each
631 step, compute the corresponding IRBS score over training data, and finally
632 find the threshold value that leads to the highest IRBS value. We then test
633 the validity of the threshold by computing its IRBS score over the verification
634 data.

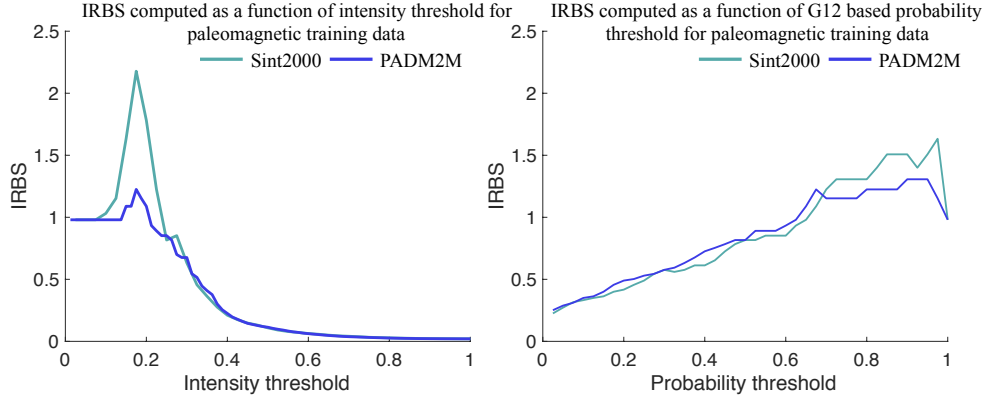


Figure 14: Determining optimal intensity and probability thresholds. Shown is IRBS over paleomagnetic training data as a function of the intensity (left) and G12-based probability (right) thresholds.

635 4.4.1. Intensity threshold-based predictions

636 An example of a threshold-based prediction strategy for dipole reversals is
 637 “a reversal will happen within the horizon if the intensity drops below a given
 638 threshold”. Note that this strategy relies on the intuitive fact that a reversal
 639 is more likely to occur in the near future if the paleomagnetic intensity is
 640 low, and that it does not make use of any dynamical considerations.

641 As explained above, we split the paleomagnetic data into two parts,
 642 “training data” and “verification data”. The training data are the signed
 643 relative paleointensity from 2 Myr to 1.05 Myr ago, which includes five rever-
 644 sals, two of which occurred close to each other to define the Cobb mountain
 645 subchron, about 1.19 Myr ago. The verification data are the signed relative
 646 paleointensity from 1.05 Myr ago onwards, and include two reversals. We
 647 apply this strategy to Sint-2000 and PADM2M and consider a 4 kyr horizon.
 648 We show thresholds and associated IRBS values over the training data in the
 649 left panel of figure 14. We observe a well-defined extremum with IRBS well
 650 above one at an intensity threshold of 0.175 for both data sets (IRBS is 2.17
 651 for Sint-2000 and 1.22 for PADM2M). This graph thus suggests that rely-
 652 ing on an intensity threshold may indeed be a meaningful way of predicting
 653 reversals within a 4kyr time-window. However, a posteriori using this op-
 654 timal intensity threshold of 0.175 fails to predict several reversals, not only
 655 within the verification data, but also within the training data. Failure to cor-

rectly predict several reversals occurs independently of whether we Sint-2000 or PADM2M (failures occurring when using Sint-2000 are illustrated in the bottom right panel of figure 15). The failure of this intensity threshold-based prediction strategy is interesting in two respects. Firstly, it shows that no intensity threshold-based strategy for either data set could pass our tests, which in turn suggests that the Earth’s dynamo may not have an intensity threshold that can be used to infer that a reversal will inevitably occur (or at least we do not have data to back up such a strategy). Secondly, the result illustrates the fact that a prediction strategy scoring IRBS well above one over the available training data may still fail to provide relevant reversal predictions, even within the training data.

Testing the same intensity threshold-based prediction strategy when considering synthetic data produced by the four low-dimensional models also leads to instructive results. The data we use are those shown in figure 2, which we again split into training and verification data. In the case of the B13 or G12 based SDE models, we find that no threshold yields IRBS larger than one, whether considering the training or even the entire data sets. This suggests that the intensity of these models can become arbitrarily low without necessarily leading to a reversal. In the case of P09, the situation is slightly different and a maximum of 1.15 can be found for IRBS when considering a threshold 0.051. However, the threshold is rather low and the corresponding maximum IRBS value is poorly defined (the graph of IRBS vs. threshold is flat and does not exhibit a distinguished global maximum). Indeed, using the optimal threshold fails to lead to a successful prediction of all reversals within training and verification data which, as before, suggests that the intensity of the P09 model can also be very low without necessarily leading to a reversal. Experiments with synthetic data of the G12 model however result in successful predictions of all reversals by this intensity threshold-based prediction strategy. We find a clear IRBS maximum of 2.64 at a threshold 0.25 over the training data, which indeed is comparable to the threshold we obtained from Sint-2000 and PADM2M (see figure 14). In this respect, G12 appears to be more Earth-like than the stochastic models. On the other hand, it appears to be more predictable by intensity threshold-based strategies than the Earth’s dynamo. This point will be further discussed below.

4.4.2. *Probability threshold-based predictions*

We now wish to test if low-dimensional models combined with data assimilation can provide a threshold criterium that is more reliable than the

693 data-derived intensity threshold above. We thus modify the above intensity
694 threshold-based strategy and predict that a reversal will occur with proba-
695 bility one within 4 kyr if the computed probability of an upcoming reversal
696 exceeds a threshold, otherwise assign probability zero.

697 We first consider probabilities derived from the G12 model. The corre-
698 sponding results are shown in the right panel of figure 14, where we show
699 IRBS for the training data as a function of the probability threshold. We ob-
700 serve that the graph flattens for probability thresholds larger than 70%, and
701 drops quickly for high probabilities larger than 98% for both paleomagnetic
702 data sets. Specifically, the optimal threshold based on Sint-2000 is 97.5%,
703 and for PADM2M threshold values between 90% – 95% are optimal, leading
704 to IRBS values of 1.63 for Sint-2000, and 1.31 for PADM2M. When these op-
705 timal thresholds are used, we obtain an IRBS of 1.13 for the verification data
706 of Sint-2000 and between 1.98 and 3.97 for the verification data of PADM2M
707 (with optimal thresholds between 90% – 95%). In addition, both reversals
708 within the verification data sets, whether Sint-2000 or the PADM2M, are
709 correctly predicted (see figure 15).

710 While the G12 probability threshold-based strategy is somewhat success-
711 ful, it also has weaknesses. For example, it leads to one false alert and fails
712 to predict the reversal ending the Cobb mountain subchron (see zoom (c) in
713 figure 15), when considering training data of Sint-2000. However, the false
714 alert precedes a reversal by only 13 kyr and the reversal is correctly predicted
715 by a later alert. In view of the much longer “typical” chron durations, such
716 a false alert may be viewed as a “slightly too early” warning. Note that
717 assessing the success of predictions by just relying on IRBS ignores the fact
718 that predicting a reversal slightly too early is an error that is less severe than
719 not predicting it at all.

720 Failing to predict the reversal ending the Cobb mountain subchron is
721 of greater concern. This reversal occurred, according to the Sint-2000 data
722 set, to within 4kyr of the previous one. Failure to predict this reversal thus
723 may result from inaccuracies within the Sint-2000 data. However, it may
724 also suggest that the G12 model is incapable of producing two successive
725 reversals within a few thousand years. Indeed, similar issues arise when using
726 the PADM2M data set. In this case, no false alert occurs before the Cobb
727 mountain subchron. However, a false alarm does occur shortly after (1kyr
728 after the subchron), again indicating some incompatibility of the G12 model
729 with this quick sequence of two reversals. The G12 model in combination with
730 PADM2M and a probability threshold-based prediction strategy further fails

731 to predict the upper Olduvai reversal (1.77 Myr ago) in the training data set.
732 In this case, the alert is triggered only once the reversal actually occurred. We
733 did not observe this behavior when using Sint-2000, which suggests that this
734 behavior may indicate the limits of probability threshold-based strategies,
735 especially in view of uncertainties in Sint-2000 or PADM2M.

736 We also apply the probability threshold-based prediction strategy to syn-
737 thetic data of the G12 model, which yields positive results. We find an
738 optimal probability-threshold of 87.5% and associated IRBS of 7.92 for the
739 verification data, as well as fully successful predictions of all reversals. These
740 tests indicate that some of the above issues could be caused by intrinsic
741 limitations of the G12 model.

742 Finally, we also test probability threshold-based strategies for the three
743 stochastic low-dimensional models P09, B13 and G12 based SDE. For the B13
744 and G12 based SDE models, no probability thresholds leading to IRBS signif-
745 icantly larger than can be found, whether considering Sint-2000, PADM2M or
746 synthetic data. This is reminiscent of the results we obtained by the intensity
747 threshold-based strategy (see above). In other words, neither B13 nor the
748 G12 based SDE model seem to provide successful probability threshold-based
749 predictions, even when considering synthetic data produced by the models.
750 The situation is again different for the P09 model. When using Sint-2000
751 data, the optimal threshold is 0.55, and the associated IRBS is 1.4 for the
752 training data, and 0.99 for the verification data. Considering PADM2M data
753 leads to a different, perhaps more encouraging result. We obtain an optimal
754 threshold of 0.275 yielding an IRBS of 1.19 for the training data, and 2.47 for
755 the verification data. However, when we consider synthetic data, we obtain
756 a lower optimal probability threshold of 0.125, leading to an IRBS of 1.96
757 for the training data, and 0.5 for the verification data, failing to success-
758 fully predict reversals. The synthetic data experiment thus suggests that the
759 probability-threshold based strategy is in fact not more applicable to P09
760 than to the other two stochastic models. These results are similar to what
761 we found when we considered intensity threshold-based predictions for the
762 P09 model (see above).

763 5. Summary and discussion

764 5.1. Summary of data assimilation

765 We considered three existing low-dimensional models, B13, P09 (both
766 stochastic, Buffett et al. (2013); Pétrélis et al. (2009)) and G12 (deterministic,

767 Gissinger (2012)), and also proposed a new scalar stochastic model, the G12
 768 based SDE, to describe the dynamics of the Earth’s magnetic dipole over
 769 geological time scales (millions of years).

- 770 1. We find that the scaling of G12 model time is limited to match either
 771 a millennium scale, or a geological time scale. While this may be an
 772 intrinsic limitation of this model, it does not prevent the G12 model
 773 from being useful in the context of the present study, provided we use
 774 the millennium time scale.
- 775 2. We calibrated all four low-dimensional models to paleomagnetic data
 776 over the past 2 Myr by using “data assimilation”. This was done by sev-
 777 eral numerical data assimilation techniques and by assimilation of two
 778 paleomagnetic data sets, Sint-2000 (Valet et al., 2005) and PADM2M
 779 (Ziegler et al., 2011).
- 780 3. We showed that all four low-dimensional models are compatible with
 781 both paleomagnetic data sets in the sense that average errors after data
 782 assimilation are no larger than 8%, provided a suitable numerical data
 783 assimilation method is used.

784 5.2. Summary of coarse reversal predictions

785 We further investigated the extent to which dipole reversals can be pre-
 786 dicted to occur within time windows of 4kyr and 8kyr, without paying at-
 787 tention to the precise timing of the reversals within the time windows. The
 788 value of such coarse predictions was assessed by hindcasting experiments,
 789 i.e., “predicting past events”, as is commonly done in numerical weather
 790 prediction. This led to the following findings.

- 791 1. Hindcasting experiments with data assimilation of synthetic data, i.e.,
 792 data produced by the models, suggest that all four models (B13, P09,
 793 G12, G12 based SDE) are intrinsically predictable for time windows of
 794 4 kyr, a necessary condition for the models to be useful as a prediction
 795 tool for Earth’s dipole. The B13, P09 and G12 based SDE models are
 796 also intrinsically predictable over 8 kyr time windows.
- 797 2. When assimilating paleomagnetic data, as documented by Sint-2000 or
 798 PADM2M, and considering 4 kyr time windows, all four low-dimensional
 799 models perform “better”, than making trivial reversal predictions based
 800 on reversal statistics of the past 2 Myr, as measured by higher inverse
 801 relative Brier scores (IRBS). Consistent with the results from synthetic

802 data experiments, the P09, B13 and G12 based SDE models also perform
803 well for 8 kyr windows. These findings suggests that low-dimensional
804 models can indeed provide “useful” information and serve as a tool to
805 understand and interpret paleomagnetic data.

- 806 3. Intensity threshold-based predictions are unsuccessful in the sense that
807 we can not obtain intensity thresholds from a “training data” set (of
808 about 1 Myr, including five reversals), that lead to success when applied
809 to a “verification data” set (of about 1 Myr, including two reversals).
810 This purely data-based strategy fails to predict several reversals in both
811 the training and verification data sets. This was found to be true for
812 Sint-2000 and PADM2M data and suggests that, given the available
813 data, paleomagnetic intensity can become low without necessarily being
814 followed by a reversal within the next 4 kyr.
- 815 4. Similar intensity threshold-based prediction tests applied to synthetic
816 data of the three stochastic models (B13, P09, G12 based SDE) suggest
817 that the intensity of these models can be low without necessarily being
818 followed by a reversal within the next 4kyr. The deterministic G12
819 model on the other hand seems to have an intensity threshold, i.e., a
820 reversal of the G12 dipole will necessarily occur if its intensity drops
821 below a threshold.
- 822 5. Probability threshold-based predictions raise an “alert” for a reversal to
823 occur within the next 4 kyr if the probability of a reversal inferred from
824 low-dimensional models and data assimilation exceeds a given thresh-
825 old. This strategy yields improved coarse predictions provided the G12
826 model is used. In contrast, stochastic models (B13, P09 and G12 based
827 SDE) give unsatisfactory results. However, even when using the G12
828 model, probability threshold-based predictions have weaknesses. These
829 are likely due to uncertainties of the Sint-2000 and PADM2M data we
830 have not properly accounted for, as well as an inability of G12 to pro-
831 duce nearby reversals. The resulting “partial” failures, however, are
832 not critical, and we conclude that a probability threshold-based strat-
833 egy using the G12 model is more reliable than a purely data-based
834 intensity threshold-based strategy.
- 835 6. Similar probability threshold-based prediction tests applied to syn-
836 thetic data from the four low-dimensional models (B13, P09, G12 and
837 G12 based SDE) further suggest that this strategy indeed fails for all
838 stochastic models (B13, P09, or G12 based SDE), but not for the de-
839 terministic G12 model. The G12 model is the only model we consider

840 for which a probability threshold can be found beyond which a reversal
841 will necessarily occur.

842 All these results taken together provide interesting evidence that determin-
843 istic low-dimensional models such as G12 in combination with data assimila-
844 tion can possibly provide a means for forecasting reversals within 4 kyr time
845 windows. It should be stressed, however, that the amount of paleomagnetic
846 data we use for these tests is limited (only 2 Myr of data, documenting only
847 seven reversals) and that errors affecting these data may not be properly
848 accounted for. The above findings should thus be interpreted with caution.

849 *5.3. Geophysical discussion and future work*

850 Assessing whether or not reversals of the geomagnetic field can be fore-
851 casted is a challenging task which has already been addressed in the past.
852 For example, several researchers have studied general characteristics of past
853 reversals as well as the behavior of the field shortly before reversals (see, e.g.,
854 Valet and Fournier, 2016, for a recent review). Others have investigated the
855 cause of the present fast decrease of the dipole field, which may be akin to
856 processes that lead to reversals (see, e.g., Hulot et al. (2002); Finlay et al.
857 (2016)). Precursors of reversals were also identified from three-dimensional
858 numerical dynamo simulations (see, e.g., Olson et al., 2009). However, iden-
859 tification of precursors within the details of the Earth’s magnetic field before
860 it reverses is difficult because of the particularly complex and varied ways
861 the field can reverse, as is documented by paleomagnetic records and three-
862 dimensional numerical simulations (see, e.g., Hulot et al., 2010a; Glatzmaier
863 and Coe, 2015). As a matter of fact, no convincing precursor has yet been
864 found in the way the modern field behaved in the recent past (see, e.g.,
865 Constable and Korte, 2006; Laj and Kissel, 2015). The search for precur-
866 sors is further limited by the fact that details of the geomagnetic field are
867 unlikely to be predictable beyond a century, as shown by investigations of
868 three-dimensional numerical dynamo simulations (Hulot et al., 2010b; Lhuil-
869 lier et al., 2011a). This limit of predictability is comparable to the time scale
870 with which the detailed morphology of the geomagnetic field changes (Hulot
871 and Le Mouél, 1994; Lhuillier et al., 2011b), but is much shorter than the time
872 elapsed between reversals. This implies that the precise timing of a reversal
873 (to within, say, a century) is likely to remain unknown until the reversal is
874 just about to happen. However, this limit does not preclude that general
875 macroscopic conditions for a reversal to occur within a wider time window

876 could be found by examining the long-term dynamic behavior of the dipole
877 field itself, which indeed displays a rich low-frequency temporal spectrum
878 (Constable and Johnson, 2005). In this context, the horizon of predictability
879 of the coarse behavior of the dipole field may be larger than that of the de-
880 tailed behavior of the full field of the Earth’s dynamo. This is the possibility
881 we investigated here with the help of data of the past behavior of the dipole
882 field, as documented by Sint-2000 and PADM2M, tentative low-dimensional
883 models of the geodynamo, and data assimilation.

884 Two key results of geophysical relevance were obtained. One is that the
885 available paleointensity data (Sint-2000 or PADM2M) do not seem to display
886 any intensity threshold below which a reversal can be guaranteed to occur
887 within the next 4 kyr. The second is that, in contrast, the very same data can
888 be assimilated by the deterministic G12 model to make reliable predictions of
889 reversals within 4 kyr time windows. It is important to emphasize that these
890 results rely on the assumption that the signed relative paleointensity data
891 provide a reliable source of information and accurately reflect the millennium
892 dynamics of the Earth’s magnetic field. Given our current understanding
893 of the way sediments record this signal, these assumptions may not hold
894 (see, e.g., Valet and Fournier, 2016, for a discussion). In particular, relative
895 timing of reversals with respect to the original paleointensity record is difficult
896 to guarantee within a few kyr, and such paleointensity data are known to
897 fail to record weak field intensities. In addition, the way sediment data
898 average the original field intensity implies that paleointensity data contain
899 some information about the near-future field intensity, at least up to 1kyr,
900 and possibly slightly beyond.

901 Another important limitation of the present study, which we already
902 stressed, is the limited amount of reversals documented in the Sint-2000
903 and PADM2M data sets. This limitation, combined with the uncertainties
904 affecting the data, may well impact IRBS, the exact values of the various
905 thresholds we computed and, therefore, the significance of our results. How-
906 ever, the consistency of our findings with respect to the data, i.e., whether
907 we use the Sint-2000 or PADM2M data sets, is encouraging.

908 Our study also revealed a number of interesting properties of the low-
909 dimensional models we considered. While all four models succeed at assimi-
910 lating the signed paleointensity data with comparable success (average errors
911 after data assimilation are no larger than 8%), and appear to be intrinsically
912 predictable in the coarse sense we defined, only predictions based on the de-
913 terministic G12 model pass the set of tests we devised. However, even the

914 G12 model may not be considered as “satisfactory” for the purpose of coarse
 915 dipole predictions. For example, it fails to properly handle fast sequences
 916 of two successive reversals (such as those bounding the Cobb mountain sub-
 917 chron). It also produces sequences that display an intensity threshold that
 918 can be used to raise successful reversal alerts for G12, contrary to the paleoin-
 919 tensity data as documented by Sint-2000 and PADM2M. Moreover, the G12
 920 model is unable to properly reproduce the observed reversal frequency when
 921 scaled to the millennium time scale. Nonetheless, the successes of the G12
 922 model in combination with the probability threshold-based prediction strat-
 923 egy indicates that these predictions may improve if “better” low-dimensional
 924 models could be obtained.

925 It is interesting in this respect to compare dipole data of the G12 model
 926 (not using any data assimilation) with the signed paleointensity data of Sint-
 927 2000 and PADM2M, and to investigate the causes of its success and failures.
 928 Comparing figures 1 and 2 (see also figure 8) makes it clear that the G12
 929 dipole data is more regular than the paleomagnetic data. The fact that an
 930 intensity threshold can be found in the case of G12, and not in the case of
 931 the paleomagnetic data, can be traced back to this regularity. Local minima
 932 that do not lead to reversals in the G12 synthetic data are all of comparable
 933 magnitude. This is not the case in the paleomagnetic data. This is also
 934 not the case in the synthetic data produced by the three stochastic models
 935 B13, P09 and G12 based SDE, which were also found to lack reliable in-
 936 tensity thresholds (with the only possible exception of P09, which however
 937 displays a very low and poorly defined intensity threshold, as described in
 938 section 4.4.1). In this respect, the dipole variable of the G12 model may
 939 be too regular when compared to Sint-2000 or PADM2M. Some regularity,
 940 however, has been found in the paleointensity data when the field approaches
 941 a reversal. In particular, it appears that this paleointensity tends to gradu-
 942 ally decrease over a period of several 10 kyr before the reversal occurs (Valet
 943 et al., 2005). This medium-term dynamics is also found in dipole data pro-
 944 duced by G12. Figure 5 compares G12 dipole data with the paleointensity
 945 data of Sint-2000 during the Brunhes-Matuyama reversal. The figure shows
 946 that the synthetic data displays a gradual decrease at a rate comparable to
 947 the average rate seen in the paleointensity data, before dropping and leading
 948 to the reversal. No similar systematic feature is found in the synthetic data
 949 produced by the three unsuccessful stochastic models. This leads us to in-
 950 terpret that the success of G12 at correctly predicting reversals is resulting
 951 from the data assimilation scheme being capable of correctly picking up this

952 trend in the paleointensity data, and thus setting G12 on its reversal path.
 953 This interpretation is also consistent with the fact that G12 partly failed at
 954 raising the proper alerts for the two reversals bounding the Cobb Mountain
 955 subchron, since the second reversal was not preceded by a medium-term in-
 956 tensity decrease. It is also consistent with the fact that G12 succeeded at
 957 forecasting reversals despite its failure to properly account for the frequency
 958 of reversals. What really matters is the sequence of events preceding the re-
 959 versal over the millennium timescale, which G12 was scaled to capture, and
 960 not the time elapsed since the last reversal.

961 The success of G12 at predicting past reversals may be a motivation to
 962 look for even better low-dimensional models, and the tests we derived provide
 963 means to assess any such model. The above discussion also highlights the
 964 fact that what matters most for a model to be a successful improvement
 965 upon G12 is that it better captures the dynamical path to a reversal. This
 966 was not the case of the three stochastic models we tested.

967 Possible routes to improvement of such stochastic models are to derive
 968 systems of SDEs (rather than scalar SDEs), as well as to include correlated
 969 noise terms (as in Buffett and Matsui (2015)). Improved deterministic models
 970 may be found as well. G12, in particular, could be improved by considering
 971 higher order terms or additional equations, e.g., more flow and field variables,
 972 while respecting the symmetries imposed by the background rotation. If the
 973 model dynamics become rich, one may need to account for the smoothing
 974 effect of sedimentation when considering the paleomagnetic data, but this
 975 could be handled, e.g., one could consider data assimilation with observation
 976 operators that model the sedimentation process. Finally, we note that $2\frac{1}{2}$ -D
 977 dynamos (e.g., Sarson and Jones, 1999) could also be tested. With modern
 978 computers, data assimilation for such models is feasible, even over geological
 979 time scales. Any improvements, however, will depend on the validity of
 980 our underlying assumption that general conditions for reversals to occur are
 981 dictated by the average large-scale behavior of the dipole field, and not by
 982 the detailed morphology of the field, which plays a role only once the reversal
 983 is just about to happen. Although our study suggests this could be the case,
 984 this still needs to be confirmed.

985 For the time being, and based on what could be achieved using the G12
 986 model and assimilating Sint-2000 and PADM2M data (up to 1kyr ago), it is
 987 reassuring to see that no warning of any reversal is currently being raised for
 988 the next few millennia by our probability threshold-based approach. This
 989 result is consistent with the fact, already pointed out by several authors

(e.g., Constable and Korte, 2006; Hulot et al., 2010a), that the current short-term fast decrease of the dipole field cannot alone be taken as evidence for an imminent reversal, even though it may possibly lead to temporarily low dipole field values (see, e.g., Laj and Kissel, 2015).

Acknowledgements

We thank Jean-Pierre Valet of the Institute de Physique du Globe de Paris and Leah Ziegler of Oregon State University for assistance in using the SINT-2000 and PADM2M datasets. We thank Bruce Buffett, University of California at Berkeley, Francois Petrélis and Christophe Gissinger of Ecole Normale Supérieure, Paris, for interesting discussion and for providing their models for this study. We also thank Olivier Sirol of IPGP for valuable help with computing.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under contract DE-AC02005CH11231, by the National Science Foundation under grants DMS-1217065 and DMS-1419044, by IPGP’s Visiting Program, CNRS PNP program, and by the Alfred P. Sloan Foundation through a Sloan Research Fellowship awarded to MM. This is IPGP contribution n° XXXX.

Appendix A. Average e -folding time of the G12 model

The e -folding time describes the time required for errors to grow by a factor e and, thus, provides a measure of how far into the future one can rely on G12 based predictions. For example, once small errors are amplified to be macroscopic, model based predictions are dominated by error. One can thus expect that G12 based predictions can be reliable at most for time-horizons comparable to the model’s e -folding time. Similarly, propagating information from data backwards in time over several e -folding times will be numerically difficult.

We estimate the e -folding time of G12 as follows. First we determine an initial condition on the attractor by simulating G12 for 10 Myr from an arbitrary point in state space; the last state of this simulation is likely to be on the attractor, or at least close to it. We pick this state to be the initial condition, and perturb it by a Gaussian with mean zero and covariance 10^{-10} times the identity matrix I . We generate 100 random perturbations and, for

each of these, compute the error as a function of time for the next 4 Myr. The error is the Euclidean norm of the difference of the reference solution and the perturbation. The average error over the 100 samples can be used to estimate the e -folding time by a log-linear least squares fit.

Our estimate of the e -folding time depends on where we start the simulations. To account for this variation, we average the e -folding time over the attractor, and repeat the above procedure with the last state of the reference trajectory serving as the initial condition for the next calculation. We do this 500 times to obtain 500 samples of the e -folding time at various locations of 2000 Myr on the attractor. The results are shown in figure A.16. We then compute the average e -folding time over these 500 samples and this average e -folding time is 40 kyr.

Appendix B. Overview of the data assimilation methods we used

The goal in data assimilation is to combine a mathematical model with information from sparse and noisy data. This is done via Bayesian statistics and conditional probability. Here we briefly review data assimilation and summarize the numerical techniques we use. More detailed reviews of data assimilation in geophysics can be found in Bocquet et al. (2010); van Leeuwen (2009); Fournier et al. (2010); Blayo et al. (2014). For earlier applications of data assimilation in geomagnetism, see Fournier et al. (2007); Sun et al. (2007); Fournier et al. (2011); Aubert and Fournier (2011); Morzfeld and Chorin (2012).

Appendix B.1. Data assimilation with deterministic models

Suppose you have a mathematical model in the form of an ordinary differential equation (ODE) (e.g., the G12 model). After discretization, e.g., with a Runge-Kutta scheme, the discrete model can be written as

$$x^n = \mathcal{M}_n(x^0),$$

where x^n is an m -dimensional column vector approximating the solution of the underlying ODE at some time t_n , and where x^0 is the state at time 0, i.e., the initial condition of the ODE. For example, for the G12 model, $x^n = [D(t = t_n), Q(t = t_n), V(t = t_n)]^T$, in which superscript T means transpose. Suppose you have collected data at time t_n . Then the state at time 0 and the data at time t_n are connected by

$$z^n = h(\mathcal{M}_{t_n}(x^0)) + v, \tag{B.1}$$

1056 where z^n is a k -dimensional vector containing the data, $h(x)$ is a given vector
1057 function, and v is a random variable that accounts for the imperfection of
1058 the mathematical model and measurement. We will assume throughout that
1059 v is Gaussian with mean zero and with a given $k \times k$ symmetric and positive
1060 definite covariance matrix R . The above equation (B.1) defines the *likeli-*
1061 *hood* $p(z^n|x^0)$, which describes the probability of the data given the initial
1062 condition x^0 . Here and below, a vertical bar denotes conditioning of random
1063 variables.

1064 We assume that the state at time 0 is not completely known, but described
1065 by a *prior* probability density $p(x^0)$, which may be a Gaussian with a given
1066 mean and variance. The prior is chosen before the data are collected. The
1067 prior and the likelihood jointly define a *posterior* probability

$$p(x^0|z^n) \propto p(x^0)p(z^n|x^0), \quad (\text{B.2})$$

1068 which contains all the information we have given the model and the data. For
1069 example, one can use the posterior distribution to compute the conditional
1070 mean, which is the minimum mean square error estimate of the state (see,
1071 e.g., Chorin and Hald (2013)).

1072 In data assimilation we find the posterior distribution by various numer-
1073 ical techniques. In the case of variational data assimilation (Bennet et al.,
1074 1993; Talagrand and Courtier, 1987), one finds the most likely state, given the
1075 data, by maximizing the posterior probability. Alternatively, Monte Carlo
1076 sampling can be used to obtain an empirical estimate of the posterior (Kalos
1077 and Whitlock, 1986; Atkins et al., 2013; Chorin and Hald, 2013). This em-
1078 pirical estimate consists of a set of weighted samples $\{w_j, X_j^0\}$, $j = 1, \dots, M$,
1079 such that averages over the samples converge to expected values with re-
1080 spect to the posterior. The Monte Carlo approach also makes it possible to
1081 incorporate errors (in model and data) into our estimation. For example,
1082 the accuracy of a state estimate can be known by computing the standard
1083 deviations of the samples. In addition, each sample can be used to produce
1084 an individual forecast, so that the Monte Carlo approach can lead to reliable
1085 forecasting, in which the uncertainty in the estimate is accounted for. In
1086 practice, many variants of these methods can be used. Below, we summarize
1087 the techniques we relied on.

1088 *Appendix B.1.1. Implicit sampling*

1089 Implicit sampling is a technique that combines ideas from variational data
1090 assimilation with Monte Carlo sampling. Details and different implementa-

1091 tions of implicit sampling can be found in Chorin and Tu (2009); Chorin
 1092 et al. (2010); Morzfeld et al. (2012); Atkins et al. (2013); Morzfeld and Chorin
 1093 (2012). Here, we only briefly describe the principle of the algorithm.

1094 The samples are generated by a data-informed probability. To find this
 1095 probability, define

$$F(x^0) = -\log p(x^0|z^n) = -\log p(x^0) - \log p(z^n|x^0).$$

1096 Specifically, for a Gaussian prior with mean μ_0 and covariance Σ_0 , and for
 1097 $v \sim \mathcal{N}(0, R)$, we find that

$$F(x^0) = \frac{1}{2} (x^0 - \mu_0)^T \Sigma_0^{-1} (x^0 - \mu_0) + \frac{1}{2} (h(\mathcal{M}_{t_n}(x^0)) - z^n)^T R^{-1} (h(\mathcal{M}_{t_n}(x^0)) - z^n).$$

1098 Let

$$\mu = \arg \min F(x^0), \quad \phi = \min F(x^0),$$

1099 be the minimizer and minimum of F , respectively, and let H be the Hessian
 1100 of F at the minimum (i.e., the $m \times m$ symmetric positive definite matrix
 1101 whose elements are the second derivatives of F). In implicit sampling, the
 1102 samples are generated by the Gaussian

$$X_j^0 \sim \mathcal{N}(\mu, H^{-1}),$$

1103 and the weights are

$$w_j \propto \exp (F_0(X_j^0) - F(X_j^0)),$$

1104 where

$$F_0(x^0) = \phi + \frac{1}{2} (x^0 - \mu)^T H (x^0 - \mu),$$

1105 is the Taylor approximation of F to second order. In summary, the implicit
 1106 sampling algorithm is:

- 1107 1. find the minimum of F (similar to variational data assimilation);
- 1108 2. generate samples using the Gaussian $\mathcal{N}(\mu, H^{-1})$;
- 1109 3. compute the weights $w_j = \exp(F_0(X_j^0) - F(X_j^0))$ for each sample.

1110 The result is a set of weighted samples which approximate the posterior
 1111 probability (B.2).

1112 *Appendix B.1.2. Sequential data assimilation*

1113 The data assimilation approach can be extended to data assimilation
 1114 problems with more than one datum. Suppose there are n data points
 1115 $z^1, \dots, z^i, \dots, z^n$, collected at times $t_1, \dots, t_i, \dots, t_n$. Then the posterior
 1116 probability (B.2) becomes

$$p(x^0|z^{1:n}) \propto p(x^0)p(z^1|x^0) \cdots p(z^i|x^0) \cdots p(z^n|x^0),$$

1117 where we use the notation $z^{1:n}$ for the set of vectors $\{z^1, \dots, z^i, \dots, z^n\}$, and
 1118 the “likelihood” of each datum, $p(z^i|x^0)$, is specified by an equation of the
 1119 form (B.1). For example, if the noise at time t_i is Gaussian with mean zero
 1120 and variance R_i , then $p(z^i|x^0) = \mathcal{N}(h(\mathcal{M}_{t_i}(x^0)), R_i)$.

1121 One can modify this approach to work sequentially as follows. Suppose
 1122 n data are available at times t_1, \dots, t_n . We first pick the first $\ell < n$ of these
 1123 data and compute the posterior

$$p_\ell(x^0|z^{1:\ell}) \propto p(x^0)p(z^1|x^0)p(z^2|x^0) \cdots p(z^\ell|x^0).$$

1124 This can be done using the same implicit sampling technique as before.
 1125 We however next remove the weights by a resampling step, during which
 1126 we delete samples with a small weight, and duplicate samples with a large
 1127 weight (see, e.g., Doucet et al. (2001) for resampling algorithms). The re-
 1128 sult is a set of M unweighted samples of this first posterior at time 0. The
 1129 samples are informed by the first ℓ data points. We then propagate these
 1130 samples forward to time t_ℓ by the model:

$$X_j^\ell = \mathcal{M}_{t_\ell}(X_j^0), \quad j = 1, \dots, M$$

1131 and compute the mean and variance of these samples to construct a Gaussian
 1132 $p(x^\ell)$ that describes the state at time t_ℓ .

1133 This Gaussian $p(x^\ell)$ is next used as a *prior* for the state at time t_ℓ , to
 1134 proceed with the assimilation of the next ℓ data points. We simply update
 1135 this prior to the posterior

$$p_l(x^\ell|z^{\ell+1:2\ell}) \propto p(x^\ell)p(z^{\ell+1}|x^\ell)p(z^{\ell+2}|x^\ell) \cdots p(z^{2\ell}|x^\ell)$$

1136 and use the same implicit sampling and resampling steps as above to draw
 1137 samples X_j^ℓ from this posterior. These unweighted samples then represent
 1138 the state at time t_ℓ , given the data $z^{1:2\ell}$. At this point, the information from

1139 the first ℓ data points was used in the prior $p(x^\ell)$, and the next ℓ data points
 1140 were used to update this prior to the posterior. These samples can then again
 1141 be forwarded, now to time $t_{2\ell}$, to produce a Gaussian *prior* $p(x^{2\ell})$ for the
 1142 state at time $t_{2\ell}$, which can again be used to proceed with the assimilation
 1143 of the next ℓ data points. This process can be repeated, using ℓ data per
 1144 sweep, until all data are assimilated. We will refer to this method as the
 1145 sequential data assimilation with implicit sampling method for deterministic
 1146 models (D-IMP, for short).

1147 *Appendix B.1.3. The ensemble Kalman filter*

1148 The ensemble Kalman filter (EnKF) is a different numerical data assimilation
 1149 technique, which computes a Gaussian approximation of the posterior
 1150 probability $p(x^n|z^{1:n})$ at any time t_n when data are collected (Evensen, 2006).
 1151 The EnKF is recursive algorithm and works as follows. First recall that z^n is
 1152 assumed to satisfy (B.1), however we assume for EnKF that the “observation
 1153 operator” h is linear, i.e., $h(x) = Hx$, where H is a matrix. Next, suppose
 1154 you have M samples of the posterior at time $n - 1$, $X_j^{n-1} \sim p(x^{n-1}|z^{1:n-1})$.
 1155 Then, for each sample, compute

$$\hat{X}_j^n = \mathcal{M}_{t_n}(X_j^{n-1}),$$

1156 and let C be the sample covariance matrix. With this covariance, define the
 1157 Kalman gain

$$K = CH^T(HCH^T + R)^{-1},$$

1158 where R is the covariance matrix of the random variable v . The Kalman gain
 1159 is used to compute the “analysis ensemble”:

$$X_j^n = \hat{X}_j^n + K \left(\hat{z}_j^n - H\hat{X}_j^n \right),$$

1160 where \hat{z}_j^n is a “perturbed observation” obtained from $\hat{z}_j^n = z^n + V_j$, V_j being
 1161 a sample of v .

1162 The EnKF then provides a state estimate at each time t_n when the data
 1163 are collected. Note that EnKF produces a Gaussian approximation of the
 1164 posterior. This can lead to large errors in nonlinear problems, where this
 1165 approximation is not valid. We will refer to this method as the EnKF method
 1166 for deterministic models (D-EnKF, for short).

1167 *Appendix B.2. Data assimilation with stochastic models*

1168 Data assimilation can also be applied to stochastic models (such as the
1169 B13 and P09 models considered in this study). It is typical in data assim-
1170 ilation to consider only discrete-time models and we follow suit. A time
1171 discretization of an SDE (1) can be written as

$$x^n = \hat{f}(x^{n-1}) + \hat{g}(x^{n-1})\Delta W,$$

1172 where \hat{f} and \hat{g} depend on the discretization we use, and where ΔW is a
1173 Gaussian with mean zero and whose variance is equal to the time step size δt
1174 (see, e.g., Kloeden and Platen (1999)). Data are collected at discrete times:

$$z^n = h(x^n) + v^n,$$

1175 where v^n are independent Gaussian random variables with mean zero and
1176 variance R^n .

1177 The posterior of interest is $p(x^{0:n}|z^{1:n})$ and a sequential approach, based
1178 on the recursion,

$$p(x^{0:n}|z^{1:n}) \propto p(x^{0:n-1}|z^{1:n-1}) p(x^n|x^{n-1})p(z^n|x^n), \quad (\text{B.3})$$

1179 is often used. Here, we use a sequential Monte Carlo approach (Doucet
1180 et al., 2001), and apply Monte Carlo sampling (recall above) at each step
1181 of the recursion to the “update” of the posterior, $p(x^n|x^{n-1})p(z^n|x^n)$. The
1182 “prior”,

$$p(x^n|x^{n-1}) = \mathcal{N}\left(\hat{f}(x^{n-1}), \delta t \hat{g}(x^{n-1})\hat{g}(x^{n-1})^T\right)$$

1183 is then defined by the discretized stochastic model, while the “likelihood”,

$$p(z^n|x^n) = \mathcal{N}(h(x^n), R^n),$$

1184 is defined by the data. The product of the prior and likelihood thus defines
1185 the posterior update we sample at each step. Again we use implicit sam-
1186 pling at each step to sample the posterior update $p(x^n|x^{n-1})p(z^n|x^n)$ (for
1187 the assimilations we perform in the manuscript, implicit sampling is in fact
1188 the optimal sampling strategy, see Morzfeld et al. (2012)). Over time, one
1189 obtains, recursively, an empirical estimate of the posterior (B.3). We will re-
1190 fer to this method as the sequential data assimilation with implicit sampling
1191 method for stochastic models (S-IMP, for short).

1192 In addition, we will also use sequential importance sampling with resam-
 1193 pling (SIR) (Doucet et al., 2001). In this method, one picks the prior as the
 1194 importance function for the posterior update at each step. The weights are
 1195 proportional to the likelihood. In short, the algorithm updates the posterior
 1196 at time $n - 1$, represented by M samples to time n as follows: (i) for each
 1197 sample, simulate the model to time n ; and (ii) compute the weight from the
 1198 likelihood $p(z^n|x^n)$; repeat for all M samples. This method is easy to imple-
 1199 ment, however becomes inefficient if the dimension of the problem increases.
 1200 We will refer to this method as the SIR method.

1201 Finally, we will also use EnKF for data assimilation with the stochastic
 1202 models. Indeed, EnKF can readily be extended to stochastic models by
 1203 generating the “forecast ensemble” (see above) with the stochastic model.
 1204 The remaining formulas of EnKF for stochastic models are then as defined
 1205 above. We will refer to this method as the EnKF method for stochastic
 1206 models (S-EnKF, for short).

1207 Amit, H., Leonhardt, R., Wicht, J., 2010. Polarity reversals from paleo-
 1208 magnetic observations and numerical dynamo simulations. *Space science*
 1209 *reviews* 155 (1-4), 293–335.

1210 Atkins, E., Morzfeld, M., Chorin, A., 2013. Implicit particle methods and
 1211 their connection with variational data assimilation. *Monthly Weather Re-*
 1212 *view* 141, 1786–1803.

1213 Aubert, J., Fournier, A., 2011. Inferring internal properties of Earth’s core
 1214 dynamics and their evolution from surface observations and a numerical
 1215 geodynamo model. *Nonlinear Processes in Geophysics* 18, 657–674.

1216 Bennet, A., Leslie, L., Hagelberg, C., Powers, P., 1993. A cyclone prediction
 1217 using a barotropic model initialized by a general inverse method. *Monthly*
 1218 *Weather Review* 121, 1714–1728.

1219 Blayo, E., Bocquet, M., Cosme, E., Cugliandolo, L. F., 2014. *Advanced Data*
 1220 *Assimilation for Geosciences*. Oxford University Press.

1221 Bocquet, M., Pires, C., Wu, L., 2010. Beyond Gaussian statistical modeling
 1222 in geophysical data assimilation. *Monthly Weather Review* 138, 2997–3023.

1223 Brendel, K., Kuipers, J., Barkema, G., Hoyng, P., 2007. An analysis of the
 1224 fluctuations of the geomagnetic dipole. *Physics of the Earth and Planetary*
 1225 *Interiors* 162, 249–255.

1226 Brier, G., 1950. Verification of forecasts expressed in terms of probability.
1227 Monthly Weather Review 78 (1), 1–3.

1228 Buffett, B., 2015. Dipole fluctuations and the duration of geomagnetic po-
1229 larity transitions. Geophysical Research Letters 42, 7444–7451.

1230 Buffett, B., Matsui, H., 2015. A power spectrum for the geomagnetic dipole
1231 moment. Earth and Planetary Science Letters 411, 20–26.

1232 Buffett, B., Ziegler, L., Constable, C., 2013. A stochastic model for paleomag-
1233 netic field variations. Geophysical Journal International 195 (1), 86–97.

1234 Buffett, B. A., King, E. M., Matsui, H., 2014. A physical interpretation of
1235 stochastic models for fluctuations in the earth’s dipole field. Geophysical
1236 Journal International 198 (1), 597–608.

1237 Cande, S., Kent, D., 1995. Revised calibration of the geomagnetic polar-
1238 ity timescale for the late cretaceous and cenozoic. Journal of Geophysical
1239 Research: Solid Earth 100, 6093–6095.

1240 Chorin, A., Hald, O., 2013. Stochastic tools in mathematics and science, 3rd
1241 Edition. Springer.

1242 Chorin, A., Morzfeld, M., Tu, X., 2010. Implicit particle filters for data
1243 assimilation. Communications in Applied Mathematics and Computational
1244 Science 5 (2), 221–240.

1245 Chorin, A., Tu, X., 2009. Implicit sampling for particle filters. Proceedings
1246 of the National Academy of Sciences 106 (41), 17249–17254.

1247 Constable, C., Johnson, C., 2005. A paleomagnetic power spectrum. Physics
1248 of Earth and Planetary Interiors 153, 61–73.

1249 Constable, C., Korte, M., 2006. Is Earth’s magnetic field reversing ? Earth
1250 and Planetary Science Letters 246, 1–16.

1251 Doucet, A., de Freitas, N., Gordon, N., 2001. Sequential Monte Carlo meth-
1252 ods in practice. Springer.

1253 Evensen, G., 2006. Data assimilation: the ensemble Kalman filter. Springer.

1254 Feingold, G., Koren, I., 2013. A model of coupled oscillators applied to the
1255 aerosol-cloud-precipitation system. *Nonlinear Processes in Geophysics* 20,
1256 1011–1021.

1257 Finlay, C. C., Aubert, J., Gillet, N., 2016. Gyre-driven decay of the earth’s
1258 magnetic dipole. *Nature Communications* (7), 10422.

1259 Fournier, A., Aubert, J., Thébault, E., 2011. Inference on core surface flow
1260 from observations and 3-D dynamo modelling. *Geophysical Journal Inter-*
1261 *national* 186, 118–136.

1262 Fournier, A., Eymin, C., Alboussière, T., 2007. A case for variational ge-
1263 omagnetic data assimilation: insights from a one-dimensional, nonlinear,
1264 and sparsely observed MHD system. *Nonlinear Processes in Geophysics*
1265 14, 163–180.

1266 Fournier, A., Hulot, G., Jault, D., Kuang, W., Tangborn, W., Gillet, N.,
1267 Canet, E., Aubert, J., Lhuillier, F., 2010. An introduction to data assim-
1268 ilation and predictability in geomagnetism. *Space Science Reviews* 155,
1269 247–291.

1270 Gallet, Y., Hulot, G., 1997. Stationary and nonstationary behavior within
1271 the geomagnetic polarity time scale. *Geophysical Research Letters* 24 (15),
1272 1875–1878.

1273 Gissinger, C., 2012. A new deterministic model for chaotic reversals. *The*
1274 *European physical Journal B* 85, 137.

1275 Gissinger, C., Dormy, E., Fauve, S., 2010. Morphology of field reversals in
1276 turbulent dynamos. *EPL (Europhysics Letters)* 90 (4), 49001.
1277 URL <http://stacks.iop.org/0295-5075/90/i=4/a=49001>

1278 Glatzmaier, G., Coe, R., 2015. Magnetic polarity reversals in the core. In:
1279 Olson, P., Schubert, G. (Eds.), *Core Dynamics*, 2nd Edition. Vol. 8 of
1280 *Treatise on Geophysics*. Elsevier, Amsterdam, Ch. 9, pp. 279–295.

1281 Glatzmaier, G. A., Roberts, P. H., 1995. A three-dimensional convective dy-
1282 namo solution with rotating and finitely conducting inner core and mantle.
1283 *Physics of the Earth and Planetary Interiors* 91, 63–75.

1284 Hoyng, P., Ossendrijver, M., Schmitt, D., 2001. The geodynamo as a bistable
1285 oscillator. *Geophysical and Astrophysical Fluid Dynamics* 94, 263–314.

1286 Hulot, G., Eymin, C., Langlais, B., Manda, M., Olsen, N., 2002. Small-scale
1287 structure of the Geodynamo inferred from Oersted and Magsat satellite
1288 data. *Nature* 416, 620–623.

1289 Hulot, G., Finlay, C. C., Constable, C. G., Olsen, N., Manda, M., 2010a.
1290 The magnetic field of planet Earth. *Space Science Reviews* 152, 159–222.

1291 Hulot, G., Le Mouél, J.-L., 1994. A statistical approach to the Earth’s main
1292 magnetic field. *Physics of the Earth and Planetary Interiors* 82, 167–183.

1293 Hulot, G., Lhuillier, F., Aubert, J., 2010b. Earth’s dynamo limit of pre-
1294 dictability. *Geophysical Research Letters* 37, L06305.

1295 Kalos, M., Whitlock, P., 1986. Monte Carlo methods, 1st Edition. Vol. 1.
1296 John Wiley & Sons.

1297 Kloeden, P. E., Platen, E., 1999. Numerical solution of stochastic differential
1298 equations. Springer.

1299 Koren, I., Feingold, G., 2011. Aerosol-cloud-precipitation system as a
1300 predator-pray problem. *Proceedings of the National Academy of Sciences*
1301 108 (30), 12227–12232.

1302 Kuipers, J., Hoyng, P., Wicht, J., Barkema, G., 2009. Analysis of the vari-
1303 ability of the axial dipole moment of a numerical dynamo model. *Physics*
1304 *of the Earth and Planetary Interiors* 173, 228–232.

1305 Laj, C., Kissel, C., 2015. An impending geomagnetic transition ? Hints from
1306 the past. *Front. Earth Sci.*, 3:61.

1307 Lhuillier, F., Aubert, J., Hulot, G., 2011a. Earth’s dynamo limit of pre-
1308 dictability controlled by magnetic dissipation. *Geophysical Journal Inter-*
1309 *national* 186, 492–508.

1310 Lhuillier, F., Fournier, A., Hulot, G., Aubert, J., 2011b. The geomagnetic
1311 secular-variation timescale in observations and numerical dynamo models.
1312 *Geophysical Research Letters* 38, L09306.

1313 Lhuillier, F., Hulot, G., Gallet, Y., 2013. Statistical properties of reversals
 1314 and chrons in numerical dynamos and implications for the geodynamo.
 1315 *Physics of the Earth and Planetary Interiors* 220, 19–36.

1316 Lowrie, W., Kent, D., 2004. Geomagnetic polarity time scale and reversal
 1317 frequency regimes. *Timescales of the paleomagnetic field* 145, 117–129.

1318 McFadden, P., Merrill, R., McElhinny, M., Lee, S., 1991. Reversals of the
 1319 Earth’s magnetic field and temporal variations of the dynamo families.
 1320 *Journal of Geophysical Research: Solid Earth* 93, 3923–3933.

1321 Meduri, D., Wicht, J., 2016. A simple stochastic model for dipole moment
 1322 fluctuations in numerical dynamo simulations. *Frontiers in Earth Science*
 1323 4 (38).

1324 Morzfeld, M., Chorin, A., 2012. Implicit particle filtering for models with par-
 1325 tial noise, and an application to geomagnetic data assimilation. *Nonlinear*
 1326 *Processes in Geophysics* 19, 365–382.

1327 Morzfeld, M., Tu, X., Atkins, E., Chorin, A., 2012. A random map implemen-
 1328 tation of implicit filters. *Journal of Computational Physics* 231, 2049–2066.

1329 Nozières, P., 1978. Reversals of Earth’s magnetic field: an attempt at a
 1330 relaxation model. *Physics of the Earth and Planetary Interiors* 17, 55–74.

1331 Olson, P., Deguen, R., Hinnov, L. A., Zhong, S., 2013. Controls on geomag-
 1332 netic reversals and core evolution by mantle convection in the phanerozoic.
 1333 *Physics of the Earth and Planetary Interiors* 214, 87–103.

1334 Olson, P., Driscoll, P., Amit, H., 2009. Dipole collapse and reversal precur-
 1335 sors in a numerical dynamo. *Physics of the Earth and Planetary Interiors*
 1336 173 (1), 121–140.

1337 Pétrélis, F., Fauve, S., 2008. Chaotic dynamics of the magnetic field gener-
 1338 ated by dynamo action in a turbulent flow. *Journal of Physics: Condensed*
 1339 *Matter* 20, 494203.

1340 Pétrélis, F., Fauve, S., Dormy, E., Valet, J.-P., 2009. Simple mechanism for
 1341 reversals of Earth’s magnetic field. *Physical Review Letters* 102, 144503.

1342 Rikitake, T., 1958. Oscillations of a system of disk dynamos. *Mathematical*
 1343 *Proceedings of the Cambridge Philosophical Society* 54, 89–105.

- 1344 Sarson, G. R., Jones, C. A., 1999. A convection driven geodynamo reversal
1345 model. *Physics of the Earth and Planetary Interiors* 111 (1), 3–20.
- 1346 Sun, Z., Tangborn, A., Kuang, W., 2007. Data assimilation in a sparsely
1347 observed one-dimensional modeled MHD system. *Nonlinear Processes in*
1348 *Geophysics* 14, 181–192.
- 1349 Talagrand, O., Courtier, P., 1987. Variational assimilation of meteorologi-
1350 cal observations with the adjoint vorticity equation. I: Theory. *Quarterly*
1351 *Journal of the Royal Meteorological Society* 113 (478), 1311–1328.
- 1352 Valet, J.-P., Fournier, A., 2016. Deciphering records of geomagnetic reversals.
1353 *Reviews of Geophysics* 54 (2), 410–446, 2015RG000506.
1354 URL <http://dx.doi.org/10.1002/2015RG000506>
- 1355 Valet, J.-P., Meynadier, L., Guyodo, Y., 2005. Geomagnetic field strength
1356 and reversal rate over the past 2 million years. *Nature* 435, 802–805.
- 1357 van Leeuwen, P., 2009. Particle filtering in geophysical systems. *Monthly*
1358 *Weather Review* 137, 4089–4144.
- 1359 Wicht, J., Meduri, D., subm. A gaussian model for simulated geomagnetic
1360 field reversals. Submitted.
1361 URL <http://arxiv.org/abs/1501.07118>
- 1362 Ziegler, L. B., Constable, C. G., Johnson, C. L., Tauxe, L., 2011. PADM2M:
1363 a penalized maximum likelihood model of the 0-2 Ma paleomagnetic axial
1364 dipole model. *Geophysical Journal International* 184 (3), 1069–1089.

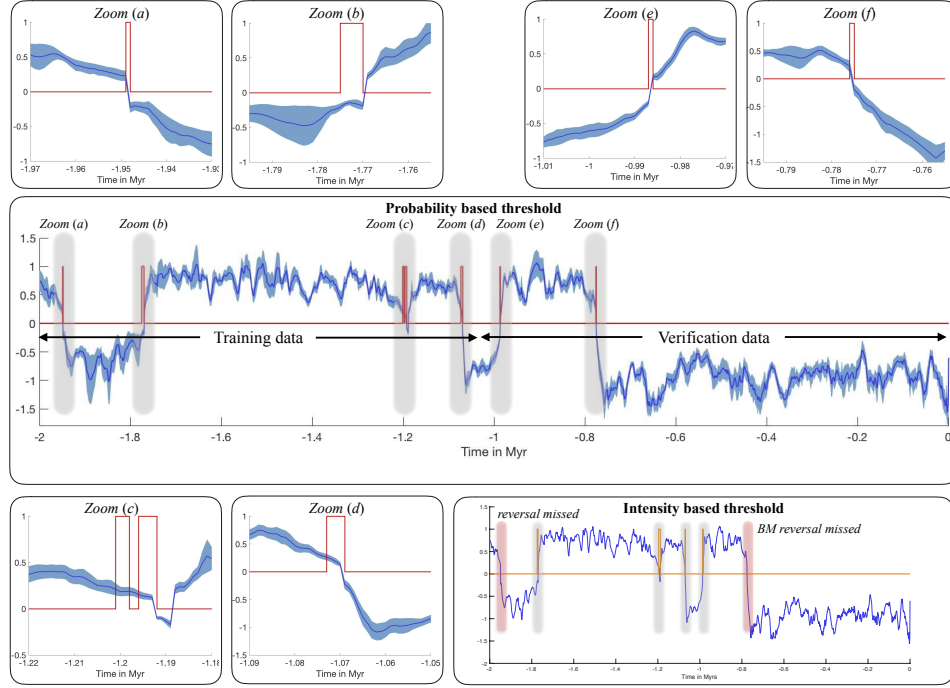


Figure 15: Illustration of probability and intensity threshold-based reversal forecasts when considering Sint-2000 data. Center panel: hindcasting by probability threshold-based strategy when relying on the G12 model; blue – Sint-2000 data; light-blue cloud – 95% confidence intervals; red – coarse reversal prediction over 4 kyr horizon (indicator function is one if a reversal is predicted to happen, zero otherwise). Top row and bottom row, left two panels: magnified data and predictions. Bottom row, right panel: hindcasting by intensity-based threshold strategy; blue – Sint-2000 data; light-blue cloud – 95% confidence intervals; orange – reversal prediction over 4 kyr horizon.

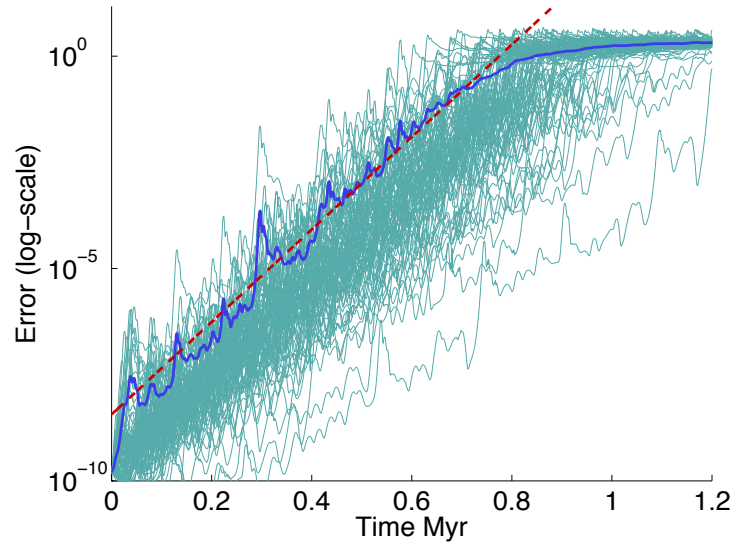


Figure A.16: Error as a function of time. The thin turquoise lines are 500 samples of the average error, each corresponding to perturbations of a given initial condition. The thick blue line is the average over these 500 samples. The red line is a log-linear fit.